AD_____

Award Number: W81XWH-12-1-0218

TITLE: Deciphering the Mechanism of Alternative Cleavage and Polyadenylation in Mantle Cell Lymphoma (MCL)

PRINCIPAL INVESTIGATOR: Chioniso Patience Masamha, Ph.D.

CONTRACTING ORGANIZATION: The University of Texas Health Science Center at Houston
Houston, TX 77030-5400

REPORT DATE: October 2014

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2014	Annual Summary	15 Sep 2013 - 14 Sep 2014

4. TITLE AND SUBTITLE

Deciphering the Mechanism of Alternative Cleavage and Polyadenylation in Mantle Cell Lymphoma (MCL)

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0218

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Chioniso Patience Masamha

E-Mail: Chioniso.P.Masamha@uth.tmc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Texas Health Science Center at Houston
Houston, TX 77030-5401

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Until recently, 3'end formation was believed to be a static event. The discovery that transformed and rapidly proliferating cells use alternative cleavage and polyadenylation (APA) to shorten the 3'UTR of their mRNAs has important implications in cancer. Truncation of the cyclin D1 mRNA in mantle cell lymphoma (MCL) is one of the earliest reported cases of APA. However, the mechanism that APA is still unknown. The goal of this project is to identify the mechanism of cyclin D1 APA regulation in cancer. So far we have been able to develop dual luciferase plasmids containing the cyclin D1 3'UTR which will enable us to determine the elements important for APA in MCL. In addition, by using RNA Seq. CFIm25 has been identified as an important global regulator of shortening of cyclin D1 mRNA and other genes. The shortened transcripts have been shown to result in increased protein levels resulting in increased cell proliferation, a hallmark of cancer. These data provides a clear link between CFIm25 and regulation of APA and the utility of using novel RNA Seq. technology. This provides a strong research platform for continued research on this project.

15. SUBJECT TERMS
Mantle cell lymphoma, alternative cleavage and polyadenylation, RNA-Seq, cyclin D1

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	31	

Table of Contents

Page

*****This second year report (2013-2014) has some sections that are identical to the first year report (2012-2013) since work done in the first year forms the basis for continued research on the project. This report will therefore include previously reported information as well as additional studies done within this past year. Please note that the additional text will be in TIMES NEW ROMAN font in order to be distinguished as new material and follows the data from the previous year.**

INTRODUCTION

The oncogene cyclin D1 is important in the disease pathogenesis of Mantle Cell Lymphoma (MCL). It has been shown that cyclin D1 undergoes alternative cleavage and polyadenylation (APA) in MCL through some, yet unknown mechanism. In order to elucidate this mechanism, this project is subdivided into 3 specific aims. The first aim is to identify the cis-elements governing cyclin D1 proximal poly (A) site selection in MCL. In addition I will determine the role of CFIm25 and CFIIm-Pcf11 in APA of cyclin D1 in MCL. The last aim involves the global analysis of genes in MCL cancer cells whose pPAS usage is regulated by CFIm-25 and CFIIm-Pcf11. This first year progress report will describe the progress made so for specific subsections of all three aims. In brief, we have successfully shown that CFIm25 plays a major role in global regulation of APA in cyclin D1 and other genes. To put the progress made in light of the specific aims into perspective, the following section will provide a brief review on the background in the field.

The role of Cyclin D1 in MCL

Overexpression of the G1-S phase cell cycle regulating oncogene cyclin D1 in MCL occurs as a result of a t(11;14)(q13;q32) chromosomal translocation that places cyclin D1 under the control of B-cell IgH genes transcription enhancers [1]. This overexpression of cyclin D1 contributes to the proliferative signature, which determines the rate of tumor proliferation and patient survival [2]. In addition to the chromosomal translocation, studies have shown that MCL patients that express cyclin D1 with a longer 3'UTR survived longer (3.28 years) than patients who expressed the truncated isoform (1.38 years)[3]. The long and the short cyclin D1 transcripts have the same open reading frame and code for the exact same protein. They only differ in the length of their 3'UTRs [3, 4]. Shortening of the cyclin D1 3'UTR in some MCL patients is due to genomic deletions. In other cases, mutations occur providing a stronger proximal PAS (pPAS) that is then used instead of the more distal PAS (dPAS) [3] through APA. However there are still cases of APA in MCL, which do not involve mutations that generate a stronger pPAS and how the pPAS is selected for use in these cases is still unknown. *Hence the working hypothesis is that in MCL cells, cyclin D1 pPAS selection is regulated by cis elements adjacent to it.* This is investigated in Aim 1.

Alternative Cleavage and Polyadenylation (APA)

Similar to what is seen in cyclin D1, most cases of alternative cleavage and polyadenylation(APA) involve using alternative polyadenylation signals (PAS) within the same terminal exon [5]. Usage of the pPAS results in shortening of the 3'UTR may result in the elimination of destabilizing elements including miRNA binding sites and AU-rich destabilizing elements (Figure 1). The resulting truncated mRNA transcripts are more stable resulting in increased translation [6]. Shortening of the 3'UTR is widespread and has been observed in rapidly proliferating cells [7] and has been implicated in tumorigenesis [6, 8]. Ongoing work in the

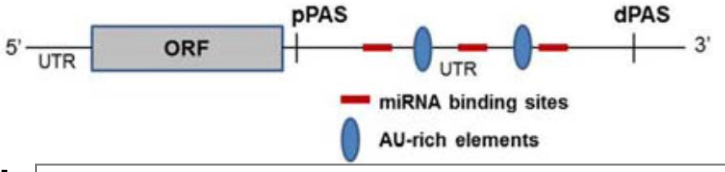


Figure 1: Use of the pPAS instead of the dPAS results in elimination of 3'UTR destabilizing elements.

field seeks to identify all the factors that regulate APA in cancer. *We hypothesized that alterations in the levels of 3'end cleavage and polyadenylation factors determine PAS site choice.* This is the goal of Aim 2.

Identification of genes in MCL that undergo APA

Since the discovery of several oncogenes that undergo APA in cancer by Mayr *et al* using Northern blots [6], next generation sequencing has been extensively used to identify APA changes on a global scale. Besides cyclin D1, the genes that undergo shortening in MCL are still unknown. Preliminary studies in our lab led us to *hypothesize that the there are other genes undergoing APA besides cyclin D1in MCL cancer cells*. This will be the focus of Aim 3.

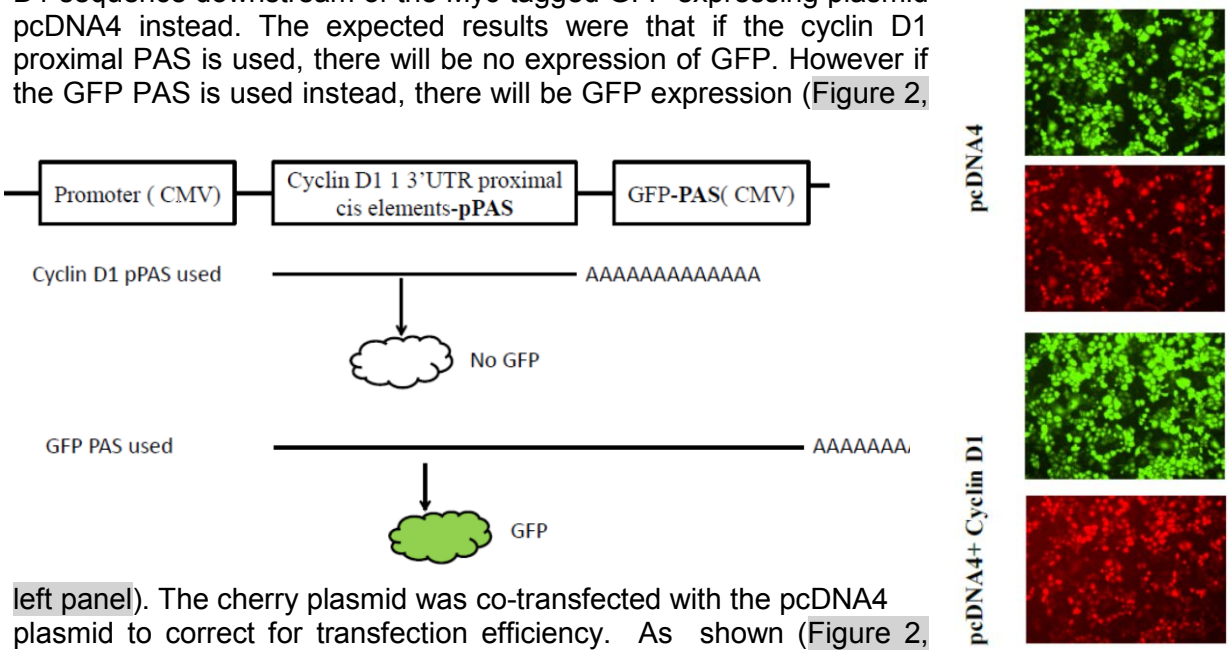
BODY

Both the negative and positive outcomes for this second year of research will be reported in this section. Research for the three aims were performed concurrently and the results will be presented for each specific aim and linked to the task listed in the Statement of Work (SOW). The progress for the second year is presented following work done in the first year to show the positive and negative outcomes for this time period relative to what was done in the first year.

Specific Aim 1: Identify the cis-elements governing cyclin D1 proximal poly (A) site selection in MCL.

Aim 1 a and b. Clone cyclin D1 upstream of EGFP in pcDNA3.1 plasmid and check for expression. Perform site directed mutagenesis to eliminate any stop codons and within Cyclin D1.

Although EGFP had already been cloned into pcDNA3.1, the plasmid did not have a tag for checking expression, so efforts were made to clone a fragment of the 3'UTR (~300nts) of cyclin D1 sequence downstream of the Myc-tagged GFP expressing plasmid pcDNA4 instead. The expected results were that if the cyclin D1 proximal PAS is used, there will be no expression of GFP. However if the GFP PAS is used instead, there will be GFP expression (Figure 2,



left panel). The cherry plasmid was co-transfected with the pcDNA4 plasmid to correct for transfection efficiency. As shown (Figure 2,

Figure 2. Left panel. The reporter system developed to detect usage of cyclin D1's proximal poly A site (pPAS). Right panel. Immunofluorescence data after cells were transfected with the pcDNA4 GFP plasmid alone or the pcDNA4 plasmid with the cyclin D1 3'UTR upstream of GFP. Either plasmid was co-transfected with cherry plasmid for use as a transfection control.

right panel) Cyclin D1 was successfully cloned in frame with GFP and we can still detect GFP expression. However, this reporter system did not attenuate GFP signal as was expected, i.e. the cyclin D1 pPAS was not used. In addition, co-transfection with cherry did not result in a reproducible immunofluorescence signal suggesting that co-transfection may alter the transfection efficiency of each plasmid.

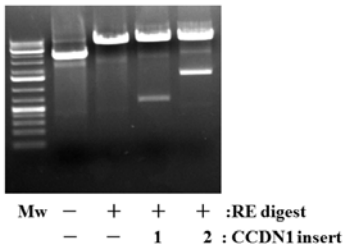


Figure 3. Restriction enzyme digest products showing the 2 different cyclin D1 3'UTR fragments cloned into psi-check.

There are a number of possibilities why the cyclin D1 pPAS was not used. It is possible that the cyclin D1 3'UTR fragment used was too short and did not include all the *cis* elements required for proper processing. To address this, 2 longer cyclin D1 3'UTR fragments (both >500bp) were generated (Figure 3) and cloned into the dual luciferase plasmid psi-check 2 (Promega). In this plasmid the cyclin D1 3'UTR plasmid fragments are cloned downstream of the human Renilla luciferase gene translational stop codon. The advantage of using this plasmid is that it also has human firefly luciferase cloned within the same reporter system allowing for intra-plasmid normalization of transfection eliminating problems observed with co-transfection of 2 different plasmids. This dual cassette reporter system is amenable to detecting effects of RNAi on cyclin D1 half-life which is one of the goals for Specific Aim 2. Since we now have the required clones, we can now proceed with site-directed mutagenesis to identify *cis*- elements and to mimic mutations observed in MCL.

Specific Aim 2: Determine the role of CFIm25 and CFIIm-Pcf11 in APA of cyclin D1 in MCL

MCL is characterized by overexpression of cyclin D1. Here we show that all the MCL cell lines have high levels of cyclin D1 which is undetectable in the normal B-cell line RPMI 1788 (Figure 4).

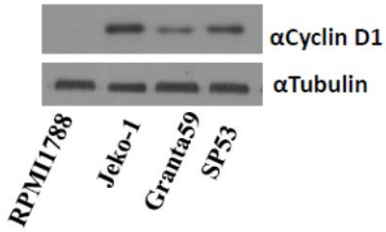


Figure 4. Western Blot analysis of normal B cell (RPMI 1788) and MCL cell lysates.

Aim 2a.

Optimize RNAi in Jeko-1 and RPMI 1788

In this aim the goal is to knock down CFIm25 and Pcf11 in RPMI 1788 normal B-cells and in the MCL cell line Jeko-1. MCL cells have been reported to be extremely difficult to transfect. In our lab we have developed a Lipofectamine based RNAi protocol that has worked in most of the cell lines we have tested so far, with a very high degree of knockdown achieved. Here we show that when we used two different siRNAs against CFIm25 in HeLa and A549 cells we were able to deplete CFIm25 levels (Figure 5a). However, we were unsuccessful in using the same technique to knockdown CFIm25 in RPMI and Jeko-1 (Figure 5b). Efforts are ongoing and in addition to use other transfection reagents we plan

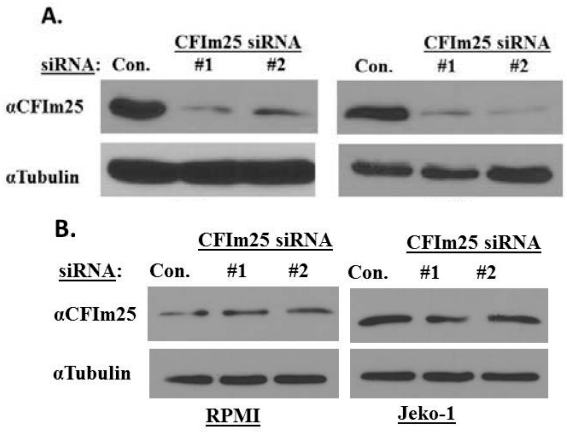


Figure 5. Western blot analysis of lysates after knocking down CFIm25 with 2 different siRNAs (#1 and #2) compared to control (Con.)siRNA.

to use Nucleofactor, which has been reported in the literature to work on MCL cells.

This Aim was giving the most problems. However, I have finally managed to successfully to perform RNAi in the B-cell lines by using electroporation and the Nucleofector kit from Lonza. The first objective was to see if I could knockdown cyclin D1 since it is highly expressed in the MCL cell lines. Cyclin D1 was successfully depleted in both the Jeko-1 and Granta cell lines (Figure 6).

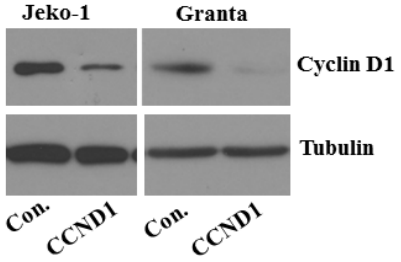


Figure 6. Western blot of Jeko-1 and Granta cell lysates after knockdown with either a control (Con.) or a Cyclin D1 (CCND1) specific siRNA.

Aim 2b. Develop RNAi resistant forms of CFIm-25 and CFIIm-Pcf11

One of the goals of this project is to use RNAi resistant plasmids to perform RNAi and rescue experiments. In collaboration with Scott Collum, a graduate student in the lab, we have developed RNAi resistant CFIm25 expression plasmid in Myc-HA-tagged pcDNA3 (Figure 7). One of the genes that underwent 3'UTR shortening after CFIm25 depletion was glutaminase (GLS) resulting in increased protein levels (Figure 7). As proof of principle, when CFIm25 siRNA was co-transfected with the CFI25m RNAi resistant plasmid, it was able to abrogate the increase in GLS protein levels (Figure 7).

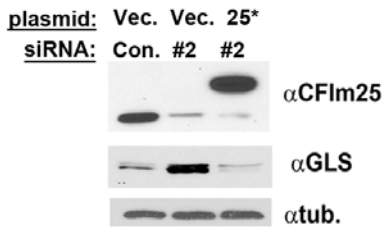


Figure 7. Western of U251 cell lysates after co-transfection of either pcDNA3 GFP (Vec) or RNAi resistant pcDNA3 CFIm25 (25*) with siRNA. The siRNA used was CFIm25 siRNA (#2) or control siRNA (Con.).

Aim 2e. Determine effects of RNAi/overexpression on cell cycle and cell proliferation

Since cyclin D1 drives the G1-S phase of the cell cycle, one of our goals is to investigate the effect of CFIm25 siRNA on cell proliferation in MCL. Preliminary data in HeLa cells shows that depletion of CFIm25 results in increased cell proliferation (Figure 8). However, the number of cells after

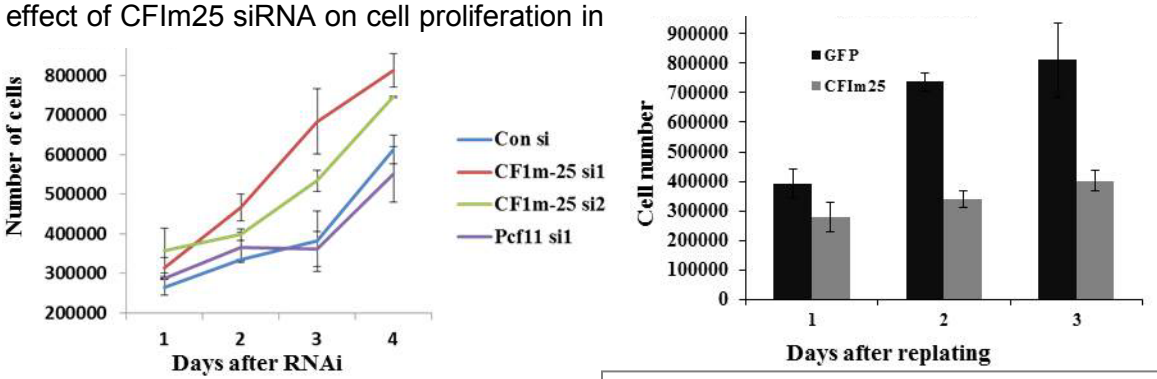
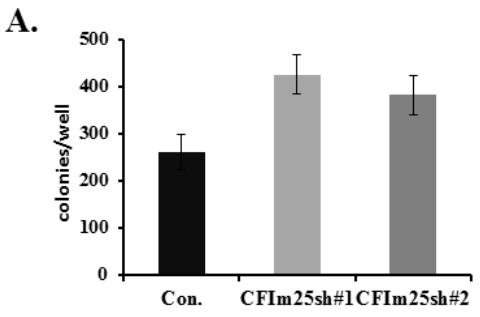


Figure 8. Left panel. Graph shows effect of different siRNA's on cell growth.
Right panel. Graphical representation of the effect of CFIm25 overexpression on cell growth.
Masamha et al,2014

depletion of Pcf11 is comparable to the control. When CFIm25 was over-expressed in U251 cells, there was a decrease in cell proliferation.

Mantle cell lymphoma patients are often diagnosed in stages III and IV with both bone-marrow involvement and leukemic spread [9]. Hence in addition to changes in cell proliferation other assays have to be used to assay for increased tumorigenicity. Although I was able to knockdown CFIm25 in MCL, the electroporation technique resulted in reduced cell numbers since some of the cells did not recover from the process. Therefore I knocked down CFIm25 in the glioblastoma cell line, LN229 (Figure 9) and performed both anchorage dependent and matrigel invasion assays. Depletion of CFIm25 resulted in decreased colony formation and increased cell invasion. Hence RNAi of CFIm25 in several different cell lines increases their tumorigenicity.



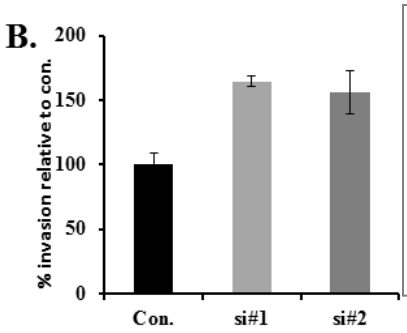


Figure 9. A. Anchorage – dependent growth assay on LN229 cells after stable knockdown of CFIm25 using shRNA #1 and #2 compared to control. B. Cell invasion assay on LN229 cells after knocking down CFIm25 using two different

Specific Aim 3: Identify genes in MCL cancer cells whose pPAS usage is regulated by CFIm-25 and CFIIm-Pcf11 by using 3'P Seq. Deep sequencing Technology

Aim 3a. Perform 3P-Seq of normal vs. MCL cancer cells

Aim 3c. Perform 3P-Seq after RNAi of cleavage factors in the MCL cell lines

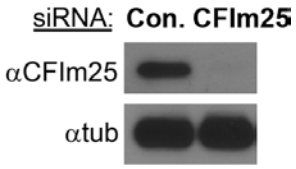


Figure 10. Western blot of HeLa cell lysates after knocking down CFIm25.

Although deep sequencing is not scheduled to be performed until the second year I will report some of the inroads we have made in our pilot studies using this technology. As a proof of principle we initially started using 3P Seq. deep sequencing technology on a sample of HeLa cells. Although we were able to identify the PAS sites used and the cleavage sites, our read density was very low i.e. less than half a million total reads. We then used another deep sequencing technique known as RNA Sequencing (RNA Seq.) which requires less mRNA partitioning and less sample manipulation. I knocked down CFIm25 in HeLa cells (Figure 10) and sent mRNA samples for sequencing at LC Sciences. This RNA-Seq. approach was very successful and resulted in ~150 million reads each for the control and the CFIm25 knockdown (Table 1)

Table 1: Summary of RNA Sequencing data

	# raw reads	# mappable reads	# total mapped reads	Percentile of mapped reads over raw reads
Control	**154,321,468**	**154,105,706**	**114,363,079**	**74.11%**
CFIm25	**146,112,668**	**145,908,590**	**118,211,444**	**80.90%**

A closer look at both the cyclin D1 and Timp2 terminal 3'UTR RNA-Seq. data (Figure 11) shows that there is a decrease in read density in the 3'UTR after CFIm25 depletion similar to previous results we obtained using qRT-PCR primers that allow us to distinguish between distal and proximal PAS usage. This RNA-Seq. data is important because: First it validates our previous qRT-PCR data that CFIm25 depletion results in 3'UTR shortening of cyclin D1 due to increased pPAS usage. Second, any changes in the 3'UTR length can be detected by analysis RNA-Seq data read density.

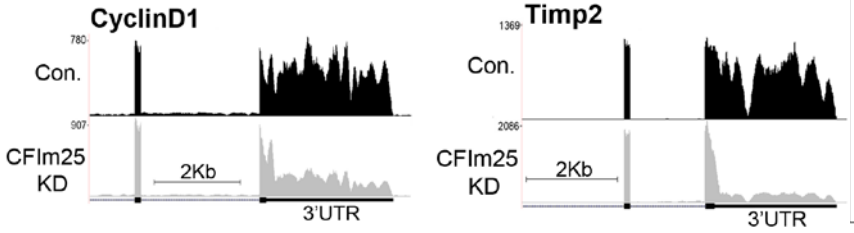


Figure 11. Analysis of the read density map of the last 2 exons of cyclin D1 and Timp-2 from RNA-Seq data. The number of reads is shown on the left for both the control (Con.) and CFIm25 siRNA (CFIm25KD).

Since analysis of the RNA-Seq data by eye is cumbersome and non-quantitative, our lab entered into a collaboration with Dr. Wei Li, a Bioinformaticist from Baylor College of Medicine whose lab specializes in developing complex algorithms to analyze genome wide sequencing data. Dr. Wei Li and his postdoctoral fellow, Dr. Zheng Xia developed a customized algorithm that is able to detect and quantify changes in 3'UTR length between samples. The change in usage of the dPAS between samples is then quantified as the percentage dPAS usage index (ΔPDUI). From this we identified 1002 transcripts that were shortened by usage of the pPAS after CFIm25 depletion, and only 5 transcripts that were lengthened (Figure 12).

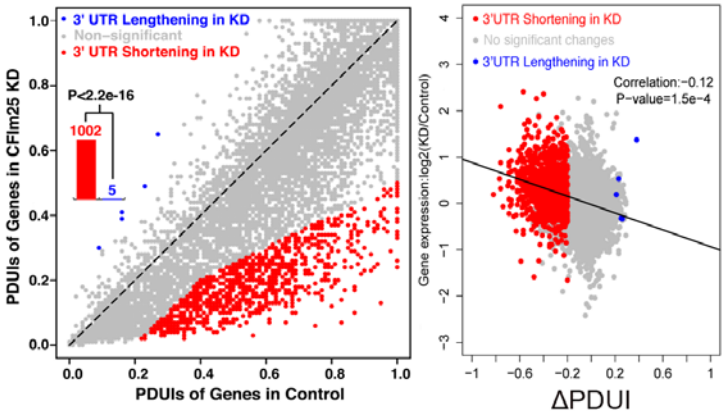


Figure 12. Left panel: Scatterplot and graph (insert) for control and CFIm25 knockdown showing changes in PAS usage on a global scale. Right panel: Shown is a plot of the correlation between usage in dPAS usage and levels of genes after CFIm25 knockdown (KD).

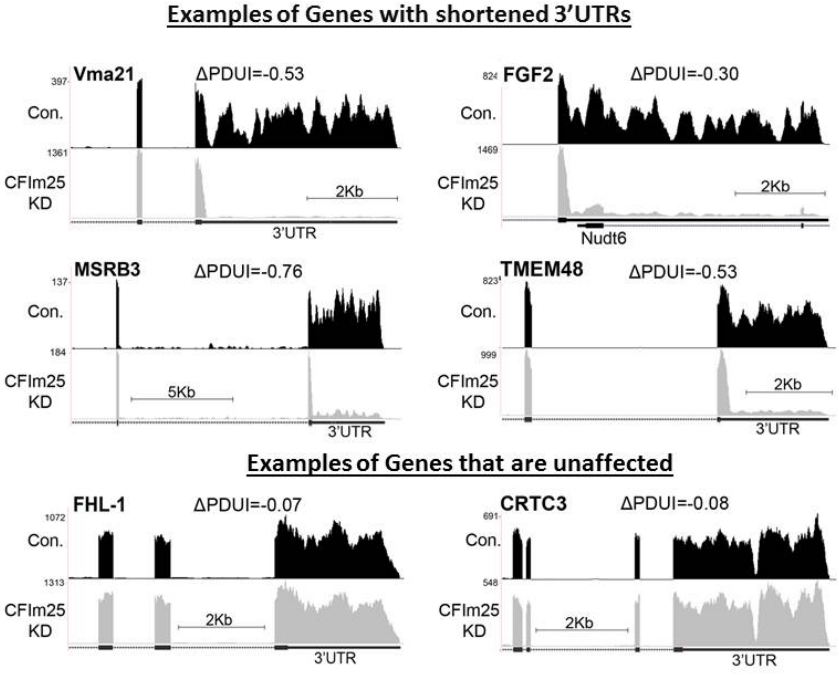


Figure 13. RNA-Seq. read densities of genes that are affected or unaffected by CFIm25 depletion.

This suggests that CFIm25 is a global repressor of pPAS usage. The algorithm also allows us to identify novel cases of APA not annotated in any database. Shown here (Figure 13) are Vma21, FGF2, MSRB3 and TMEM48 which undergo APA after CFIm25 depletion. There are also some genes that are unaffected by loss of CFIm25 despite having long 3'UTRs and the two examples shown are FHL-1 and CRTC3.

Despite successfully performing RNA-Seq. on HeLa cells last year, we had to redo the RNA-Seq. analysis in duplicate to validate that the results were reproducible according to the ENCODE Consortium Standards, Guidelines and Best Practices for RNA-Seq. We decided to perform this in the HeLa cells so we could have several (n=3) biological replicates, and compare the RNA-Seq. results from one sample that was run months before. Below (Table 2) is a summary of the RNA-Seq. data performed in duplicate [10]. There were two control replicates (Con. R1 and Con. R2), and the duplicate CFIm25 knockout (CFIm25 KDR1 and CFIm25KD R2). Interestingly we managed to get almost double the number of reads as we obtained with our trial sample (Table 2 vs.Table 1), increasing our level of confidence in the sequencing process.

Table 2: Summary of RNA-Sequencing data (Table from Masamha et al,2014)

Samples	Con. R1	Con. R2	CFIm25KDR1	CFIm25KD R2
Mapped reads	269,463,902	270,426,980	265,270,708	282,404,060
#CDS	52.4%	47.6%	53.0%	51.5%
#5' UTR	5.2%	5.2%	5.5%	5.4%
#3' UTR	27.0%	23.4%	21.4%	20.3%
#Intron	6.3%	5.8%	6.0%	5.6%
#Intergenic	9.1%	17.9%	14.1%	17.3%
3' UTR average coverage	236	205	184	185

As a proof of principle, we analyzed the read density map of several genes from all three RNA-Seq runs side by side to test for reproducibility. Shown are the read density maps from three genes(Vma21, SMOC1 and FHL1) and they are all consistently reproducible (Figure 14).

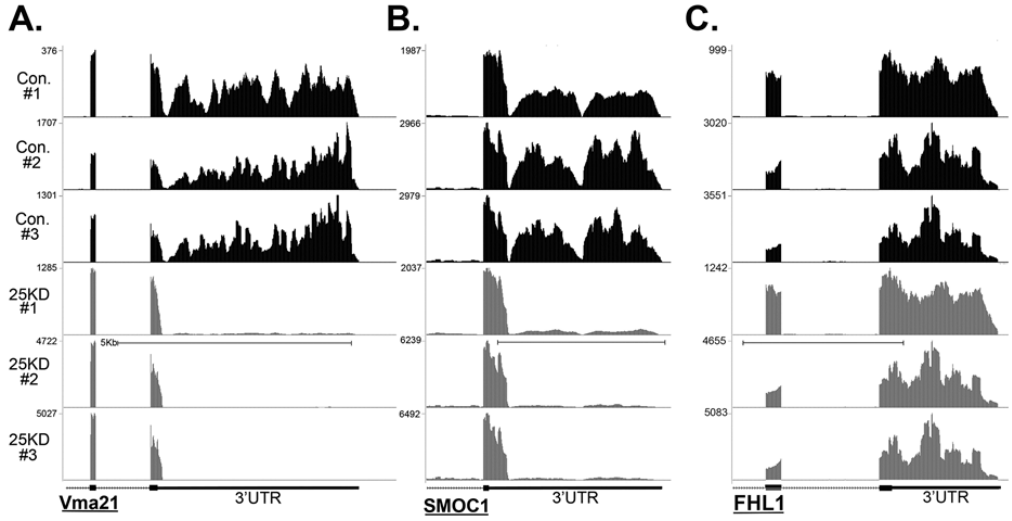


Figure 14. Read density map for from triplicate RNA-Seq. runs. There are three siRNA control samples (Con. #1-3) and three CFIm25 siRNA (25KD#1-3) A. Shows the density for the 2 last exons of Vma21 which is affected by CFIm25 depletion. B. This is the read density of the terminal exon of SMOC1 which is affected by CFIm25 depletion. C. This is an example of one gene, FHL1, which is totally unaffected by CFIm25 depletion and shown are the read density of its final two exons. (Figure from Masamha et al,2014)

Furthermore, re-analysis of the two replicates using our previously developed algorithm resulted in an increase in the number of genes from that had their 3`UTR shortened from 1,002 (Figure 12) to 1,450 after CFIm25 depletion (Figure 15) due to the increase in the RNA reads (Table 2)

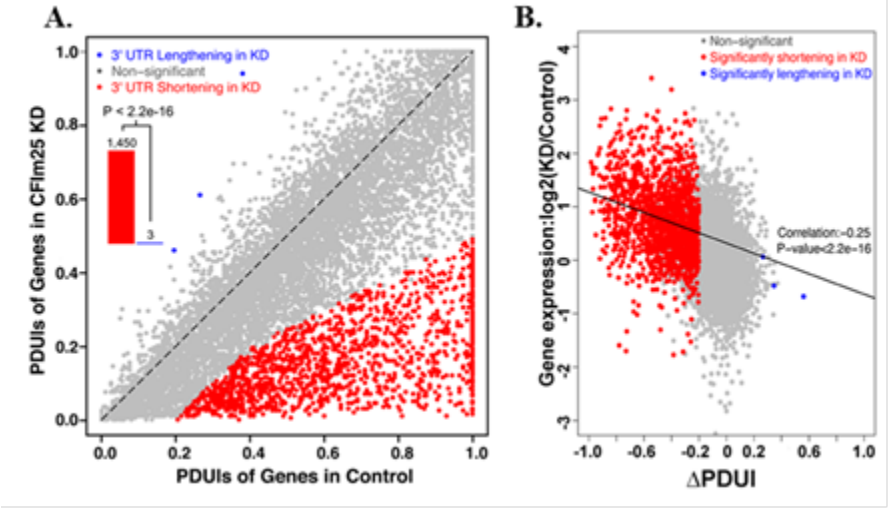


Figure 15. A: Scatterplot representation with graph (insert) for control and CFIm25 knockdown showing global changes in PAS. B: Graph showing correlation between dPAS usage and levels of gene expression after CFIm25 knockdown (KD). (Figure from Masamha et al,2014)

Since one of the goals of this aim was to identify the genes that are regulated by CFIm25 and Pcf11 in MCL cells I had to first determine the levels of each of these two factors in the different cell lines.

When compared to the normal B-cell line RPMI1788, the levels of CFIm25 are significantly decreased in Jeko-1 and Mino and slightly decreased in Granta and SP53. There are reduced levels of Pcf11 in RPMI 1788 compared to three of the MCL cell lines Jeko, Granta and SP53. Hence, both these factors may therefore be involved in regulating APA in MCL.

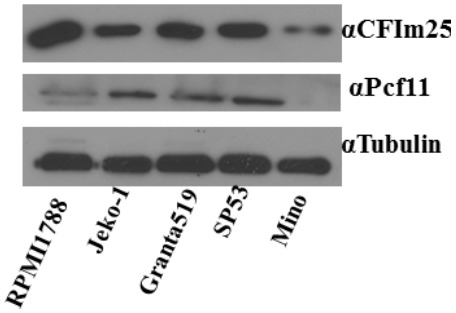


Figure 16. Western blot analysis of protein lysates from different MCL cell lines and normal B-cell line (RPMI1788).

Aim 3d. Validate 3P-Seq results on a gene by gene basis using qRT-PCR and map PAS used by 3'RACE.

To validate the results from RNA-Seq. amplicons were designed (Figure 17) for a few select genes to detect changes in distal pPAS usage normalized to total mRNA for each gene.

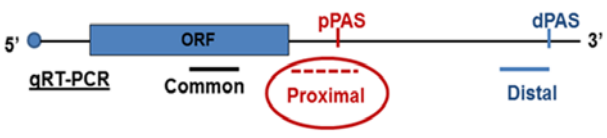
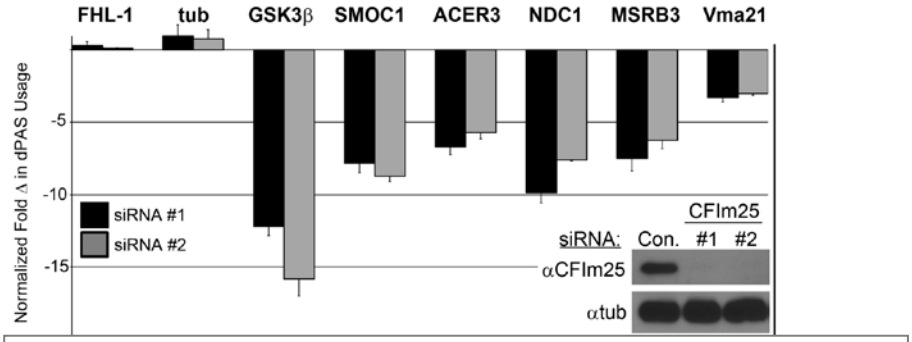


Figure 17. Location of amplicons (common to measure total levels of transcript and distal to detect changes in dPAS usage) used in qRT-PCR.

Quantitative RT-PCR (qRT-PCR) was used to detect APA after CFIm25 depletion using 2



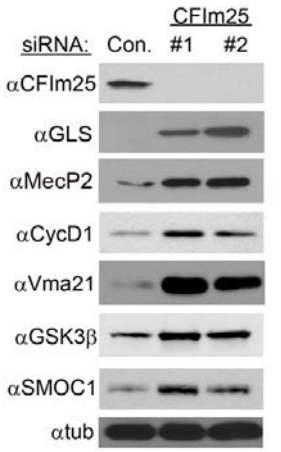


Figure 18. Left panel:qRT-PCR data showing changes in dPAS usage after CFIm25 depletion :Insert. Western blot after CFIm25 depletion using 2 different siRNAs. Right panel. Western probing for different genes after CFIm25 RNAi.

different siRNAs. As shown (Figure 18 , left panel), qRT-PCR was able to validate RNA.Seq. data of 2 genes that were unaltered by CFIm25 depletion and several genes that switched from dPAS to pPAS usage after CFIm25 knockdown. Western blot analysis was then performed to test whether this increase in pPAS usage correlates with increased protein expression. There were significant increase in protein levels for several genes which correlates with a switch to the pPAS resulting in the generation of more stable transcripts. Most notably there were increases seen in levels of glutaminase (GLS), MecP2 and cyclin D1 (Figure 18, right panel) which have been directly associated with increased cell proliferation characteristic of the transformed cell phenotype. This also correlates with increased cell proliferation, which was observed in Figure 8. Most notably, we see changes in APA correlating with increased levels of oncogenic cyclin D1, a main driver of MCL pathogenesis.

KEY RESEARCH ACCOMPLISHMENTS (2013-2014)

- CFIm25 is a global regulator of APA for over 1,400 genes including cyclin D1
- The changes from dPAS to pPAS usage regulated by CFIm25 depletion corresponds with enhanced tumorigenicity as shown by increased cell proliferation, cell invasion and anchorage dependent growth
- We have identified electroporation as a viable technique to use to perform RNAi in MCL
- We have also validated that our RNA-Seq. analysis data is highly reproducible by redoing RNA-Seq. in duplicate.

REPORTABLE OUTCOMES (2013-2014 fiscal year)

Talks:

Chioniso Patience Masamha, Zheng Xia, Jingxuan Yang, Todd Albrecht, Scott Collum, Min Li, Wei Li, Ann-Bin Shyu and Eric J. Wagner. Evading miRNA Regulation through Alternative Polyadenylation in Glioblastoma. (Selected for Plenary talk). Symposia on Cancer research, 2014. Illuminating Genomic Dark Matter "ncRNA in Disease and Cancer". The University of Texas MD Anderson Cancer Center, Houston, TX. October 9-10.

Manuscripts:

Chioniso P. Masamha*, Zheng Xia*, Jingxuan Yang, Todd R. Albrecht, Min Li, Ann-Bin Shyu, Wei Li & Eric J. Wagner. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature (2014) 510, 412–416 ; doi:10.1038/nature13261. PMID: 24814343
*First co-authors.

Training:

(Proposed in Aim 3b. Attend class at CSHL on Deep Sequencing Technology)

- Attended CSHL training course in Advanced Sequencing Technology and Applications, November 12-24, 2013.

CONCLUSIONS

This second year research progress report builds upon work done in the first year. Successfully knocking down cyclin D1 in MCL by electroporation opens an avenue that will allow us to perform further RNAi experiments in these cells. Differences in levels of CFIm25 and Pcf11 between the MCL cell lines and Pcf11lends further credence to our hypothesis that these two factors may be involved in regulating APA in MCL. We will choose one cell MCL cell line and compare it to the normal B-cell line for the RNA-Seq. experiments planned in MCL. Furthermore, the reproducibility of the RNA-Seq. data in HeLa and increased tumorigenicity observed after CFIm25 RNAi in HeLa will act as the basis for our continued proposed MCL based research.

REFERENCES

1. Bertoni, F., et al., *Update on the molecular biology of mantle cell lymphoma.* Hematol Oncol, 2006. **24**(1): p. 22-7.
2. Rosenwald, A., et al., *The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma.* Cancer Cell, 2003. **3**(2): p. 185-97.
3. Wiestner, A., et al., *Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival.* Blood, 2007. **109**(11): p. 4599-606.
4. Chen, R.W., et al., *Truncation in CCND1 mRNA alters miR-16-1 regulation in mantle cell lymphoma.* Blood, 2008. **112**(3): p. 822-9.
5. Neilson, J. and R. Sandberg, *Heterogeneity in mammalian RNA 3' end formation.* Exp Cell Res, 2011. **316**(8): p. 1357-64.
6. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells.* Cell, 2009. **138**(4): p. 673-84.
7. Sandberg, R., et al., *Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites.* Science, 2008. **320**(5883): p. 1643-7.
8. Elkon, R., et al., *E2F mediates enhanced alternative polyadenylation in proliferation.* Genome Biology, 2012. **13**(7): p. R59.
9. Pileri, S.A. and B. Falini, *Mantle cell lymphoma*. Vol. 94. 2009. 1488-1492.
10. Masamha, C.P., et al., *CFIm25 links alternative polyadenylation to glioblastoma tumour suppression.* Nature,2014. **510**, 412–416 ; doi:10.1038/nature13261. PMID: 24814343

LETTER

doi:10.1038/nature13261

CFIm25 links alternative polyadenylation to glioblastoma tumour suppression

Chioniso P. Masamha[1]*, Zheng Xia[2]*, Jingxuan Yang[3], Todd R. Albrecht[1], Min Li[3], Ann-Bin Shyu[1], Wei Li[2] & Eric J. Wagner[1]

The global shortening of messenger RNAs through alternative polyadenylation (APA) that occurs during enhanced cellular proliferation represents an important, yet poorly understood mechanism of regulated gene expression[1,2]. The 3′ untranslated region (UTR) truncation of growth-promoting mRNA transcripts that relieves intrinsic microRNA- and AU-rich-element-mediated repression has been observed to correlate with cellular transformation[3]; however, the importance to tumorigenicity of RNA 3′-end-processing factors that potentially govern APA is unknown. Here we identify CFIm25 as a broad repressor of proximal poly(A) site usage that, when depleted, increases cell proliferation. Applying a regression model on standard RNA-sequencing data for novel APA events, we identified at least 1,450 genes with shortened 3′ UTRs after CFIm25 knockdown, representing 11% of significantly expressed mRNAs in human cells. Marked increases in the expression of several known oncogenes, including cyclin D1, are observed as a consequence of CFIm25 depletion. Importantly, we identified a subset of CFIm25-regulated APA genes with shortened 3′ UTRs in glioblastoma tumours that have reduced CFIm25 expression. Downregulation of CFIm25 expression in glioblastoma cells enhances their tumorigenic properties and increases tumour size, whereas CFIm25 overexpression reduces these properties and inhibits tumour growth. These findings identify a pivotal role of CFIm25 in governing APA and reveal a previously unknown connection between CFIm25 and glioblastoma tumorigenicity.

Recently, it has become increasingly clear that mRNA 3′-end formation is subject to dynamic regulation under diverse physiological conditions[2–5]. Over 50% of human genes have multiple polyadenylation signals, thereby increasing the potential diversity in mRNA transcript length[6]. The formation of mRNA transcripts using these distinct poly(A) sites (PASs) is carried out by APA, with the most common form involving differential use of alternative PASs located within the same terminal exon (reviewed in ref. 7). Processing at the PAS most proximal to the stop codon (pPAS) removes negative regulatory elements that reduce mRNA stability or impair translation efficiency, such as AU-rich elements (AREs)[8] and microRNA (miRNA) targeting sites[9,10]. It has been reported that both rapidly proliferating cells[1,2] and transformed cells[3,11] preferentially express mRNAs with shortened 3′ UTRs. Despite these observations, the mechanisms that control the extensive distal-to-proximal PAS switch observed in proliferative and/or transformed cells, the relationship between cause and effect, and the critical target genes subject to this regulation, are not well characterized.

To measure relative changes in endogenous APA events, we devised a quantitative polymerase chain reaction after reverse transcription (qRT–PCR) assay to monitor the transcript-specific use of the distal PAS (dPAS) while normalizing for total mRNA levels for three test transcripts, cyclin D1 (*CCND1*), *DICER1* and *TIMP2*, known to undergo APA[3,12]. Using this approach, we readily detected appreciable usage of dPASs for all three genes in HeLa cells (Extended Data Fig. 1). This was somewhat surprising given their highly transformed state, but is consistent with previous reports that not all transformed cells tested exhibit appreciable 3′ UTR shortening[1,3]. Previous studies implicate multiple members of the cleavage and polyadenylation (CPA) machinery as potentially regulating poly(A) site selection[12–15]. To test the relative contribution of these factors to the APA of the three test genes, we used systematic RNA interference (RNAi) (Fig. 1a–c). We observed only small changes in the relative use of the dPAS after knockdown of members of the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CSTF) and cleavage factor IIm (CFIIm) complexes (Fig. 1d–f). By contrast, we detected significant reduction in dPAS usage after knockdown of the members of the CFIm complex. These results are consistent with a recent report that CFIm68 depletion decreases 3′ UTR length[14]; however, the most notable PAS switching was found to occur after knockdown of CFIm25. We therefore focused all further analyses on CFIm25.

Traditional methods of global PAS profiling use mRNA partitioning and digestion to sequence poly(A) junctions within messages[1,16,17]. To identify global targets of CFIm25 with a more streamlined approach requiring less sample manipulation, we performed high-depth ($>3 \times 10^8$ reads) RNA sequencing (RNA-seq) after knocking down CFIm25 in parallel with a control knockdown. We determined that 23% of RNA-seq reads can be uniquely mapped to 3′ UTRs of expressed genes leading to approximately 200-fold sequence coverage (Extended Data Fig. 2a, b). We first analysed the three test genes and observed markedly reduced read density within the 3′ UTRs in response to CFIm25 depletion (Fig. 2a). These results not only confirm our qRT–PCR findings that HeLa cells robustly use the dPAS for all three test genes under basal conditions but also demonstrate that considerable 3′ UTR shortening induced by CFIm25 knockdown is readily visualized by analysing the read density of RNA-seq data.

On the basis of this promising observation, we applied a novel bioinformatics algorithm termed 'dynamic analysis of alternative polyadenylation from RNA-seq' (DaPars; see Methods) for the *de novo* identification of all instances of 3′ UTR alterations between control and CFIm25 knockdown cells, regardless of a pre-annotated pPAS within each RefSeq transcript. DaPars uses a linear regression model to identify the exact location of this novel proximal 3′ UTR as the optimal fitting point (Fig. 2b, red point) as well as the abundance of both novel and annotated UTRs. The degree of difference of 3′ UTR usage between the samples was then quantified as a change in percentage dPAS usage index (ΔPDUI), which is capable of identifying lengthening (positive index) or shortening (negative index) within the 3′ UTR. When applied to the 12,273 RefSeq transcripts whose average terminal exon sequence coverage is more than 30-fold, DaPars identified 1,453 transcripts possessing a significant, reproducible shift in 3′ UTR usage in response to CFIm25 depletion (Fig. 2c and Extended Data Fig. 2c, d). Notably, among this group of transcripts, 1,450 are shifted to pPAS usage in CFIm25 knockdown cells. We found a significant enrichment of the CFIm25 UGUA binding motif and previously reported CFIm25 iCLIP sequence tags[14] within 3′ UTRs that shortened after CFIm25 knockdown relative to transcripts exhibiting no length change (Extended Data Fig. 3).

[1]Department of Biochemistry and Molecular Biology, The University of Texas Medical School at Houston, Houston, Texas 77030, USA. [2]Division of Biostatistics, Dan L Duncan Cancer Center and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, 77030 Texas, USA. [3]The Vivian L. Smith Department of Neurosurgery, The University of Texas Medical School at Houston, Houston, Texas 77030, USA.
*These authors contributed equally to this work.

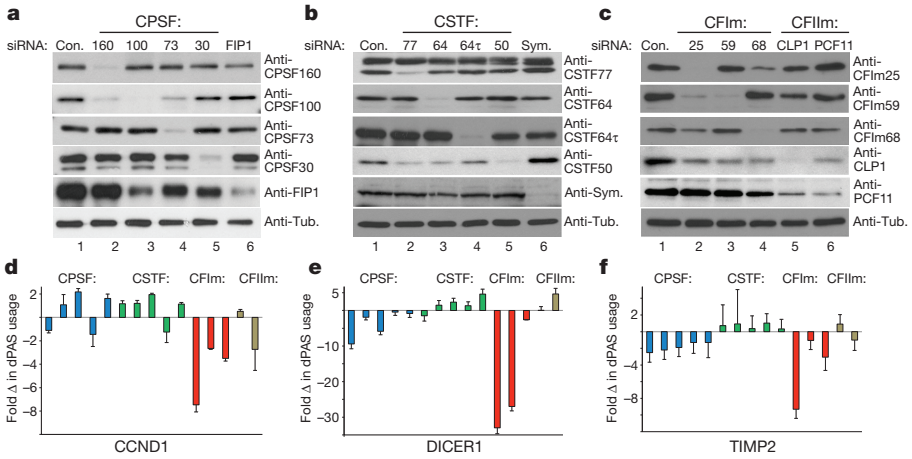


Figure 1 | CFIm25 depletion leads to consistent and robust 3′ UTR shortening of test genes. **a**–**c**, Western blot analysis of HeLa cell lysates treated with control siRNA (Con.) and siRNAs individually targeting each of members of the CPA machinery and Symplekin (Sym.). In all cases, tubulin (Tub.) was used as a loading control. **d**–**f**, Quantified results of three biologically independent qRT–PCR experiments on RNA isolated from cells represented in panels **a**–**c** with the factors presented in the same order as shown in western blots **a**–**c**. See Methods for quantification details.

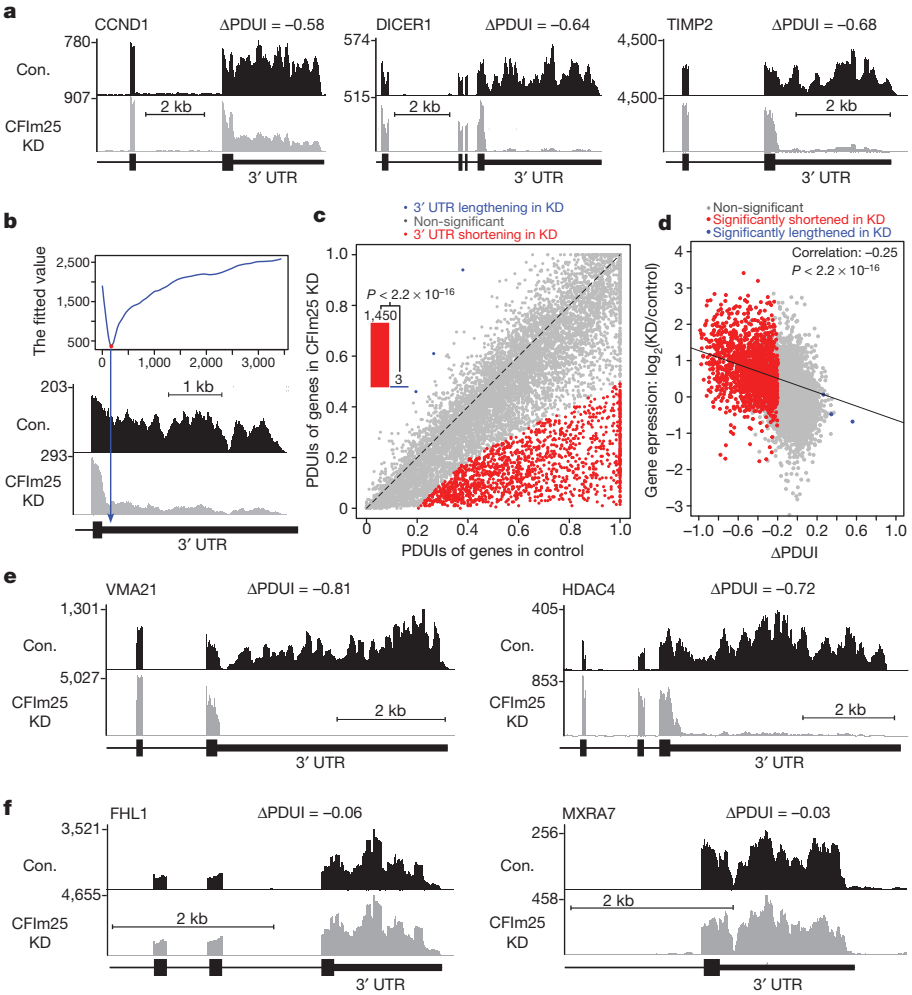


Figure 2 | The DaPars algorithm identifies broad targets of CFIm25 in standard RNA-seq data. **a**, RNA-seq read density for 3′ UTR, terminal exon and upstream exon(s) after the control (Con.) siRNA treatment and CFIm25 knockdown (KD) in HeLa cells. Numbers on *y*-axis indicate RNA-seq read coverage. **b**, Diagram depicts how the differential alternative 3′ UTR usage was identified based on DaPars. The *y*-axis shows the fitting value of the DaPars regression model and the locus with minimum fitting value (red point) is the predicted alternative pPAS for the RNA-seq data (bottom). **c**, Scatterplot of PDUIs in control and CFIm25 knockdown cells where mRNAs significantly shortened ($n = 1{,}450$) or lengthened ($n = 3$) after CFIm25 knockdown (false discovery rate (FDR) ≤ 0.05, absolute ΔPDUI ≥ 0.2 and at least twofold change of PDUIs between CFIm25 knockdown and control cells) are coloured. The shifting towards pPAS is significant ($P < 2.2 \times 10^{-16}$, binomial test). **d**, Correlation between dPAS site usage and gene expression levels of control and CFIm25 knockdown cells. The *x*-axis shows ΔPDUI; a negative value indicates that pPAS is prone to be used in CFIm25 knockdown cells. The *y*-axis shows the logarithm of the expression level of genes from the CFIm25 knockdown relative to the control sample. **e**, Representative RNA-seq density plots along with ΔPDUI values for genes whose 3′ UTR is shortened in response to CFIm25 knockdown. Numbers on *y*-axis indicate RNA-seq read coverage. **f**, Representative RNA-seq density plots along with ΔPDUI values of genes whose 3′ UTR is unchanged by CFIm25 knockdown.

Moreover, we determined that 70% of transcripts whose 3′ UTR is shortened after CFIm25 knockdown use a pPAS within the first one-third of their 3′ UTR. By contrast, only 29% of multi-PAS transcripts that did not alter 3′ UTR length in response to CFIm25 have an annotated pPAS in the first third of their 3′ UTR. This demonstrates that CFIm25 APA targets are enriched with pPASs positioned close to the stop codon to maximize their degree of 3′ UTR shortening. Collectively, these results clearly indicate that the function of CFIm25 is to broadly repress proximal poly(A) site choice, and consequently, the shortening of 3′ UTR length is considerable for the majority of CFIm25-regulated transcripts upon its depletion.

One potential consequence of 3′ UTR shortening in CFIm25 knockdown is the loss of miRNA-binding sites and/or AREs, resulting in truncated mRNA transcripts that evade negative regulation. Although the correlation between transcript expression change and ΔPDUI was modest (Pearson correlation = −0.25), it does reveal that transcripts with shorter 3′ UTR in CFIm25 knockdown cells have overall higher expression levels (Fig. 2d). We observed that 64% of transcripts with shortened 3′ UTRs exhibited significantly increased steady-state levels, 34% were unchanged, and only 2% were significantly reduced (Extended Data Fig. 4). We have also organized the list of CFIm25-regulated genes with respect to their ΔPDUI score, change in relative levels of transcript, and predicted numbers of ARE motifs and miRNA target sites lost after APA (Supplementary Table 1) and observed that gene expression positively correlates with the number of lost ARE motifs and miRNA target sites (Extended Data Fig. 5). Several examples of novel genes whose APA is regulated by CFIm25 are shown in Fig. 2e and it is important to note that not all long 3′ UTRs were observed to shorten in response to CFIm25 knockdown, indicating that the CFIm complex regulates many, but not all genes capable of APA (Fig. 2f). Collectively, these data demonstrate the power and ease of the DaPars algorithm to identify APA within standard RNA-seq, and indicate that the major form of CFIm25 regulation is to repress pPAS choice at a global level.

To validate the ΔPDUI results, we created qRT–PCR amplicons to monitor dPAS usage of six genes whose 3′ UTRs were found to be shortened after CFIm25 knockdown and two that were not altered. Using these amplicons, we analysed RNA isolated from cells effectively depleted of CFIm25 using two independent short interfering RNAs (siRNAs) (Fig. 3a, inset), and observed high congruence between qRT–PCR results and those obtained using RNA-seq and ΔPDUI (Fig. 3a, graph). To test formally for the presence of de-repressed protein expression from mRNAs with shortened 3′ UTRs, we measured their levels in lysates from knockdown cells (Fig. 3b). We observed considerable increases in protein levels of CFIm25 target genes, including several that have a well-documented role in tumour growth, such as cyclin D1, glutaminase and methyl-CpG-binding protein 2 (MECP2)[18–22]. It is worth noting that the 3′ UTR of each of these genes has been shown to be subject to miRNA-mediated inhibition[23–25]. Consistent with this observation, we also noted enhanced cellular proliferation in response to knockdown of CFIm25 relative to control knockdown in HeLa cells (Fig. 3c). Finally, to determine whether the 3′ UTR is sufficient to elicit translational de-repression of a heterologous protein in response to CFIm25 knockdown, we used reporters with the SMOC1 3′ UTR cloned downstream of luciferase or the GAPDH 3′ UTR, which was not found to alter its poly(A) site usage. We observed that only the luciferase activity specifically resulting from the luciferase–SMOC1 reporter was increased in response to knockdown of CFIm25 (Fig. 3d), supporting the idea that the increased expression of endogenous SMOC1 protein when CFIm25 is depleted is mediated through its 3′ UTR.

The collective observations that CFIm25 depletion leads to broad 3′ UTR shortening, enhanced expression of growth promoting genes and increased cell proliferation support the hypothesis that CFIm25 is a novel anti-proliferative gene whose levels may be reduced in human cancers. We focused our analysis on glioblastoma, as recent reports indicate that brain tissue possesses the longest 3′ UTRs[26,27]. We reasoned that tumours derived from these cells might be more sensitive to changes in CFIm25 levels than other cancers. To test this prediction, we downloaded

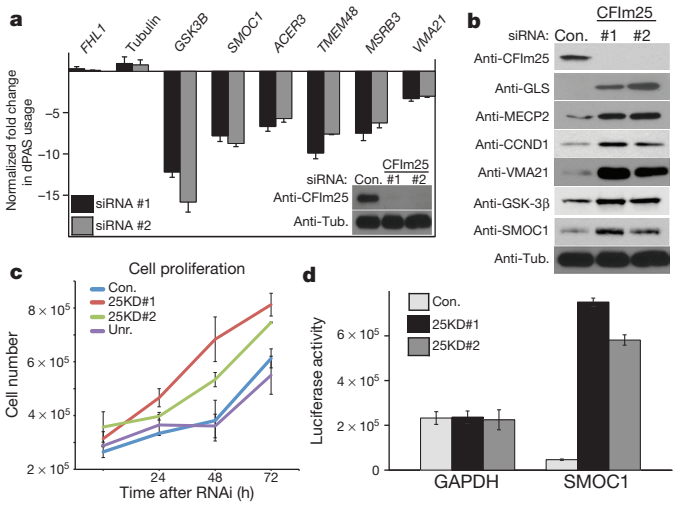


Figure 3 | Increased pPAS usage after CFIm25 depletion results in increased protein translation and enhanced cell proliferation. **a**, qRT–PCR results of select genes shown as fold change in dPAS usage after CFIm25 depletion. Experiments were performed in triplicate with data shown as mean + standard deviation from the mean (s.d.). The inset shows western blot analysis demonstrating effective knockdown of CFIm25 using two distinct siRNAs. Tub., tubulin. **b**, Results of western blot analysis of cell lysates after knockdown of CFIm25 using siRNA. **c**, Growth of HeLa cells was measured after RNAi of CFIm25 compared with cells transfected with control siRNA or the siRNA to the CFIIm complex subunit PCF11 (Unr.). Results shown are mean ± standard deviation (s.d.) ($n = 3$). **d**, Graph representing luciferase activity from cells transfected with a luciferase reporter containing the 3′ UTR of either GAPDH or of SMOC1 after being transfected with either control or CFIm25 siRNA. Data are the average of three independent experiments and error bars show s.d.

archived patient RNA-seq data from The Cancer Genome Atlas (TCGA), stratified it according to CFIm25 expression, and analysed it using DaPars. Indeed, following the same cut-offs in our HeLa RNA-seq 3′ UTR analysis, we identified 60 genes with altered 3′ UTRs, with 59 of those experiencing shortening in glioblastoma expressing lower levels of CFIm25 (Fig. 4a and Supplementary Table 2). Among those genes, a significant number of events (24 genes; $P = 2.2 \times 10^{-12}$ by hypergeometric testing) were also shortened in CFIm25 knockdown HeLa cells and this percentage of overlap increased markedly to 86% as the ΔPDUI cut-off increased from 0.2 to 0.4 (Extended Data Fig. 6). Two representative examples of genes, FOS-related antigen 2 (*FRA2*; also known as *FOSL2*) and *MECP2*, with shortened 3′ UTRs in low CFIm25-expressing glioblastoma tumours is shown in Fig. 4b, demonstrating a compelling similarity between the patient samples and HeLa cells before and after CFIm25 knockdown. Overexpression of either of these genes has been shown to enhance cell proliferation[18,28].

To test formally whether altering CFIm25 expression can modulate glioblastoma tumorigenic properties, we screened a panel of glioblastoma cell lines and observed that U251 cells naturally express lower levels of CFIm25 compared with LN229 cells (Fig. 4c). To raise CFIm25 levels in U251 cells, we created cell lines stably expressing either Myc-tagged CFIm25 or green fluorescent protein (GFP) as a control. In parallel, we used RNAi to reduce CFIm25 levels in LN229 cells (Fig. 4c). We observed a significant reduction in anchorage-dependent growth and cellular invasion in U251 cells overexpressing CFIm25 compared with the GFP control, whereas reducing CFIm25 in LN229 cells caused an increase in both of these properties (Extended Data Fig. 7). To determine if the altered *in vitro* properties of glioblastoma cells affected tumour growth kinetics *in vivo*, we used a subcutaneous xenograft model. Increased expression of CFIm25 in U251 cells resulted in a marked reduction in tumour growth and decreased tumour cell proliferation (Fig. 4d and Extended Data Fig. 8). By contrast, depletion of CFIm25 in LN229 cells caused a profound increase in tumour size (Fig. 4e and Extended

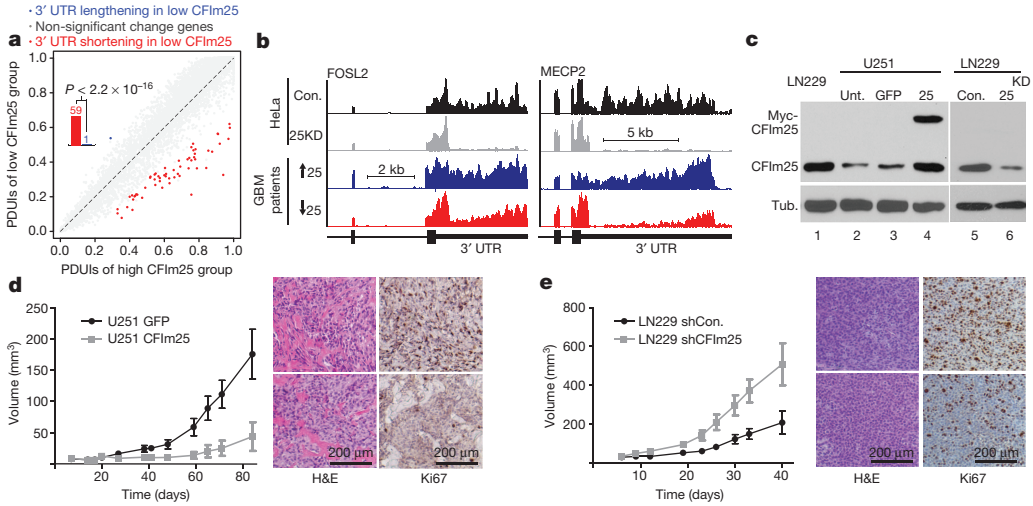


Figure 4 | Altered expression of CFIm25 modulates glioblastoma tumour growth. **a**, The global analysis of 3′ UTR changes in glioblastoma (GBM) patient samples with either high or low levels of CFIm25. Scatterplot of PDUIs from both data sets using the same cut-offs as in Fig. 2c. The shifting to pPAS in the low CFIm25 group is significant ($P < 2.2 \times 10^{-16}$; binomial test). **b**, Representative UCSC Genome Browser images of RNA-seq data, demonstrating 3′ UTR shortening after CFIm25 knockdown in HeLa cells and in glioblastoma patient samples with high (blue) or low CFIm25 expression (red). KD, knockdown. **c**, Western blot analysis of lysates from two glioblastoma cell lines. Note that the overexpressed Myc–CFIm25 also increases endogenous CFIm25 levels in U251 cells. Tub., tubulin; Unt., untreated. **d**, Growth comparison of U251 tumours overexpressing either GFP (control) or CFIm25. Data represent the average of ten mice per group. Right panel shows representative haematoxylin and eosin (H&E) and Ki67 staining of U251 GFP tumours (top) or U251 CFIm25 tumours (bottom). Scale bars, 200 μm. **e**, Growth comparison of LN229 tumours derived from cells transduced with lentiviruses expressing a scrambled short hairpin RNA (shRNA) (control) or with lentiviruses expressing shRNA targeting CFIm25. Data represent the average of ten mice per group. Right panel shows representative H&E and Ki67 staining of LN229 tumours expressing shRNA targeting CFIm25 (top) or LN229 tumours expressing scrambled shRNA (bottom). Scale bars, 200 μm.

Data Fig. 9). Collectively, these results uncover a tumour suppressive property of CFIm25 in glioblastoma that is probably mediated through its broad repression of APA-dependent mRNA 3′ UTR shortening.

We identified CFIm25 among 15 cleavage and polyadenylation factors as a key factor that broadly regulates APA. Importantly, the data presented here also extend our understanding of APA in regulated gene expression through the demonstration that extensive shortening of 3′ UTRs causally leads to enhanced cellular proliferation and tumorigenicity, probably through the upregulation of growth promoting factors, such as cyclin D1. These results indicate the importance of 3′ UTR usage in cell growth control and underscore the need for further research into the mechanism and regulation of APA and its potential links to other human diseases.

METHODS SUMMARY

Human cell lines used were cultured using standard techniques. RNAi and western blot experiments were conducted as described previously[29]. For luciferase experiments, one day after the second siRNA hit, cells were transfected with 3′ UTR *Renilla* luciferase plasmids and activity was assayed after 24 h. Total RNA for pRT–PCR was reverse transcribed using MMLV-RT (Invitrogen). qRT–PCR reactions were performed using SYBRGREEN (Fermentas). Duplicate control and CFIm25 knockdown samples were sequenced by HiSeq 2000. RNA-seq reads were aligned (hg19) using TopHat 2.0.10[30]. All the TCGA glioblastoma RNA-seq BAM files were downloaded from the UCSC Cancer Genomics Hub (https://cghub.ucsc.edu/). DaPars was used to identify differential 3′ UTR usage from RNA-seq (Z.X. *et al.*, unpublished observations; https://code.google.com/p/dapars). For tumour xenografts, U251 cells were stably transfected with GFP or CFIm25 plasmids. LN229 cells were transfected with lentivirus expressing CFIm25 shRNA. After subcutaneous injection of cell lines into nude mice, glioblastoma tumour size was monitored and tumours were removed and histologically analysed.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 January 2013; accepted 13 March 2014.
Published online 11 May 2014.

1. Elkon, R. *et al.* E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biol.* **13,** R59 (2012).
2. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320,** 1643–1647 (2008).
3. Mayr, C. & Bartel, D. P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138,** 673–684 (2009).
4. Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. USA* **106,** 7028–7033 (2009).
5. Mangone, M. *et al.* The landscape of *C. elegans* 3′UTRs. *Science* **329,** 432–435 (2010).
6. Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33,** 201–212 (2005).
7. Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* **43,** 853–866 (2011).
8. Chen, C.-Y. A. & Shyu, A.-B. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.* **20,** 465–470 (1995).
9. Farh, K. K. *et al.* The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310,** 1817–1821 (2005).
10. Wu, L. & Belasco, J. G. Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Mol. Cell* **29,** 1–7 (2008).
11. Singh, P. *et al.* Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res.* **69,** 9422–9430 (2009).
12. Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A. & Handa, H. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3′-UTRs. *Nucleic Acids Res.* **34,** 6264–6271 (2006).
13. Yao, C. *et al.* Transcriptome-wide analyses of CstF64–RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl Acad. Sci. USA* **109,** 18773–18778 (2012).
14. Martin, G., Gruber, A. R., Keller, W. & Zavolan, M. Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′UTR length. *Cell Reports* **1,** 753–763 (2012).
15. Thomas, P. E. *et al.* Genome-wide control of polyadenylation site choice by CPSF30 in *Arabidopsis*. *Plant Cell* **24,** 4376–4388 (2012).
16. Jan, C. H., Friedman, R. C., Ruby, J. G. & Bartel, D. P. Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs. *Nature* **469,** 97–101 (2011).
17. Shepard, P. J. *et al.* Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17,** 761–772 (2011).
18. Bernard, D. *et al.* The methyl-CpG-binding protein MECP2 is required for prostate cancer cell growth. *Oncogene* **25,** 1358–1366 (2006).
19. Sicinski, P. *et al.* Cyclin D1 provides a link between development and oncogenesis in the retina and breast. *Cell* **82,** 621–630 (1995).
20. Weinstat-Saslow, D. *et al.* Overexpression of cyclin D1 mRNA distinguishes invasive and *in situ* breast carcinomas from non-malignant lesions. *Nature Med.* **1,** 1257–1260 (1995).

21. Liu, W. *et al.* Reprogramming of proline and glutamine metabolism contributes to the proliferative and metabolic responses regulated by oncogenic transcription factor c-MYC. *Proc. Natl Acad. Sci. USA* **109,** 8983–8988 (2012).
22. Wang, J. B. *et al.* Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. *Cancer Cell* **18,** 207–219 (2010).
23. Klein, M. E. *et al.* Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nature Neurosci.* **10,** 1513–1514 (2007).
24. Deshpande, A. *et al.* 3′UTR mediated regulation of the cyclin D1 proto-oncogene. *Cell Cycle* **8,** 3592–3600 (2009).
25. Gao, P. *et al.* c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* **458,** 762–765 (2009).
26. Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O. & Lai, E. C. Widespread and extensive lengthening of 3′ UTRs in the mammalian brain. *Genome Res.* **23,** 812–825 (2013).
27. Ulitsky, I. *et al.* Extensive alternative polyadenylation during zebrafish development. *Genome Res.* **22,** 2054–2066 (2012).
28. Nakayama, T. *et al.* Aberrant expression of Fra-2 promotes CCR4 expression and cell proliferation in adult T-cell leukemia. *Oncogene* **27,** 3221–3232 (2008).
29. Wagner, E. J. & Garcia-Blanco, M. A. RNAi-mediated PTB depletion leads to enhanced exon definition. *Mol. Cell* **10,** 943–949 (2002).
30. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank members of the E.J.W., A.-B.S. and W.L. laboratories for helpful discussions, P. Carpenter for reviewing the manuscript, and Q. Zhu and T. Shan of LC Sciences for their efforts on the RNA-seq. This work was supported by a CPRIT grant to E.J.W. and A.-B.S. (RP100107), and in part by a Department of Defense grant to E.J.W. (W81XWH-11-1-0304), National Institutes of Health (NIH) grants to A.-B.S. (GM046454) and E.J.W. (CA167752 and CA166274), and an endowment from Houston Endowment (to A.-B.S.). Work in the W.L. laboratory is funded by grants from the Department of Defense (W81XWH-10-1-0501), CPRIT (RP110471-C3) and NIH (R01HG007538). Work in the M.L. laboratory is funded by grants from the Dr Marnie Rose Foundation and the William and Ella Owens Medical Research Foundation. C.P.M. acknowledges a Department of Defense Postdoctoral Visionary Award (W81XWH-12-1-0218).

Author Contributions A.-B.S., W.L. and E.J.W. designed the study. C.P.M., T.R.A. and J.Y. performed the described experiments with conceptual advice from M.L. W.L. and Z.X. conducted bioinformatic analyses and developed the DaPars algorithm. C.P.M., Z.X., W.L., A.-B.S. and E.J.W. wrote the manuscript.

Author Information Raw sequence data has been deposited in the Gene Expression Omnibus under accession number GSE42420. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.J.W. (Eric.J.Wagner@uth.tmc.edu), A.-B.S. (Ann-Bin.Shyu@uth.tmc.edu) or W.L. (WL1@bcm.edu).

METHODS

RNA-seq. We used whole transcriptome RNA-seq to investigate alternative PAS usage in a genome-wide fashion. Two control and two CFIm25 knockdown samples were sequenced by HiSeq 2000 (LC Sciences). Paired-end RNA-seq reads with 101 bp in each end were aligned to the human genome (hg19) using TopHat 2.0.10[30]. RefSeq gene expressions were quantified by RSEM[31]. A statistical summary of read alignments and average gene expressions can be found in Extended Data Fig. 2. More than 12,000 (~50%) human RefSeq genes can be detected through RNA-seq with expression levels more than 1 fragments per kilobase of transcript sequence per million mapped paired-end reads (FPKM)[32]. More importantly, the average of 23% of RNA-seq reads can be uniquely mapped to 3′ UTRs of expressed genes that renders around 200× coverage on UTRs. All the TCGA glioblastoma RNA-seq BAM files were downloaded from the UCSC Cancer Genomics Hub (CGHub; https://cghub.ucsc.edu/).

Analysis of APA from RNA-seq. We used a novel bioinformatics algorithm DaPars (Z.X. *et al.*, unpublished observations; https://code.google.com/p/dapars) for the *de novo* identification of APA from RNA-seq. The observed sequence coverage was represented as a linear combination of novel and annotated 3′ UTRs. For each RefSeq transcript with annotated PAS, we used a regression model to infer the end point of alternative novel PAS within this 3′ UTR at single nucleotide resolution, by minimizing the deviation between the observed read coverage and the expected read coverage based on a two-PAS model, in both control and CFIm25 knockdown samples simultaneously.

To quantify the relative PAS usage, we defined the percentage of dPAS usage for each sample as PDUI index. The greater the PDUI is, the more the dPAS of a transcript is used and vice versa.

ΔPDUI. We used the following three criteria to detect the most significant shifted 3′ UTR events: First, given the expression levels of short and long 3′ UTRs in two samples in each condition, we compute the significance of the difference of mean PDUIs using Fisher's exact test, which is further adjusted by Benjamini–Hochberg (BH) procedure to control the FDR at a level of 5%. Second, the absolute difference of mean PDUIs must be no less than 0.2. Third, the absolute $\log_2$ ratio (fold change) of mean PDUIs must be no less than 1. To avoid false positive estimation on low coverage transcripts, we required that there be more than 30-fold coverage on the 3′ UTR region of both samples. For genes with multiple annotated PASs, we only kept the one with the greatest absolute ΔPDUI value. Last, we identified 1,453 transcripts possessing a significant shift in 3′ UTR usage in response to CFIm25 knockdown, the vast majority of which have shortened 3′ UTRs in CFIm25 knockdown.

Bioinformatic analyses of 3′ UTR shortening. As miRNA binding sites and other regulatory sequences such as AREs reside in 3′ UTRs[33,34], APA has an important role in mRNA stability, translation and translocation. Indeed, it has been reported that shorter 3′ UTRs produce higher levels of protein[3]. To elucidate the consequences of 3′ UTR shortening, we provided the numbers of lost ARE motifs and miRNA binding sites due to the 3′ UTR shortening for the transcripts shifting to proximal 3′ UTR usage in CFIm25 knockdown cells (Supplementary Table 1). The ARE is one of the most prominent *cis*-acting regulatory elements found in 3′ UTRs to target mRNAs for rapid degradation[35]. The eight different consensus ARE motifs, including the plain AUUUA pentamer, were retrieved from the ARE site database[35]. miRNA–mRNA binding information was based on miRNA target prediction database TargetScanHuman version 6.2[36–38]. To limit the miRNA to high-confidence sites, we required the probability of the preferentially conserved targeting (PCT) score to be more than 0 for all highly conserved miRNA families[38].

Differentially expressed gene expression analysis. With two replicates in each group, we used edgeR[39] to call differentially expressed genes with FDR < 0.05. To better quantify gene expression with shorter 3′ UTRs, we counted reads based on the coding regions of each transcript.

Cell culture and cell counts. All the cell lines used (HeLa, U251 and LN229) were cultured in DMEM supplemented with 10% FBS (+1% penicillin and streptomycin) in a 5% CO_2 incubator at 37 °C. Cell counts were done using a standard hemacytometer.

siRNA and western blot assays. Both siRNA transfection and western blot analysis were performed as previously described[29]. The siRNA was purchased from Sigma and all the siRNAs used are shown below. After transfection, cells were harvested for mRNA extraction, western blotting or Matrigel assay. To detect 3′-end-processing factors by western blotting, the following primary antibodies from Bethyl Laboratories were used: CPSF160, CPSF100, CPSF73, CPSF30, FIP1, CSTF77, CSTF64τ, CSTF50, CFIm68 and CFIm59. Other antibodies used include CFIm25 (PTGlabs), CSTF64 and CFIIm PCF11, and Symplekin (Sigma and CFIIm CLP1 (Epitomics). Additional antibodies include VMA21, GLS, ACER3 and GSK-3β (PTGlabs); cyclin D1 (Cell Signaling); and SMOC1 and tubulin (Abcam).

siRNA sequences. We used the following siRNA sequences. CPSF160 si1: 5′-GCUUUAAGAAGGUCCCUCA; si2: 5′-CUUACCACGUGGAGUCUAA; CPSF100 si1: 5′-CUCAACUUCUUGAUCAGAU; si2: 5′-GGAUAGAUGGUGUCUUAGA; CPSF73 si1: 5′-CCAUAUACUGGUCCCUUUA; si2: 5′-GAUAUUGGAAGUUCAGUCA; CPSF30 si1: 5′-GUGCCUAUAUCUGUGAUUU; si2: 5′-CCUAUAUCUGUGAUUUGAA; FIP1 si1: 5′-CGAAUGGGACUUGAAGUUA; si2: 5′-GACAAGUACUGCCUCCAGA; CSTF77 si1: 5′-GAAGACUUAUGAACGCCUU; si2 5′-CACAGAAUCAACCUAUAGA; CSTF64 si1: 5′-GGCUUUAGUCCCGGGCAGA; si2: 5′-GGUUAUGGCUUCUGUGAAU; CSTF64τ si1: 5′-GUCUUAGAGACACGUGUAA; si2: 5′-CUAAUGUUCUGCUGAACCA; CSTF50 si1: 5′-GUCGUAAGUCCGUGCACCA; si2: 5′-CUACUCUUCGCCUUUAUGA; Symplekin si1: 5′-CAGUUCAACUCGGGCCUGA; si2: 5′-GAGACAUUGAGUUGCUGCU; CFIm25 si1: 5′-CCUCUUACCAAUUAUACUU; si2: 5′-GCUAUAUACAGUGUAGAAU; CFIm59 si1: 5′-CUCAUCUGCUCGUGUGGAU; si2: 5′-GCAAUUUCCAGCAGUGCCA; CFIm68 si1: 5′-CUGCAAUUUCUUUAAUUAA; si2: 5′-GGAUCAAGACGUGAACGAU; CFIIm CLP1 si1: 5′-GCUUAUGUCUCCAAGGACA; si2: 5′-CAGUUCAGUUGGAGUUGUU; CFIIm PCF11 si1: 5′-GUACCUUAUGGAUUCUAUU; si2: 5′-GUAUCUCACUGCCUUUACU) and the control siRNA used was described elsewhere[29].

qRT–PCR. After appropriate transfections, total RNA was extracted using TRIzol Reagent (Life Technologies) using the manufacturer's protocol. For qRT–PCR the mRNA was reverse transcribed using MMLV-RT (Invitrogen) using the manufacturer's protocol to generate cDNA. The qRT–PCR reactions were performed using Stratagene MxPro3000P (Agilent Technologies) and SYBRGREEN (Fermentas). Common primers were designed to target the open reading frame and normalize for total transcript. The distal primers were designed to target sequences just before the dPAS and detect long transcripts that use the dPAS. All primers used are shown below. Data were calculated using a modified version of the $2^{-\Delta\Delta CT}$ method to show changes in dPAS usage, where CT is the threshold cycle. First, the CT values for the common and distal amplicons were normalized to the levels of 7SK, where ΔCT (common or distal) $= CT_{\text{common or distal}} - CT_{7SK}$. Then $\Delta\Delta CT = \Delta CT_{\text{distal}} - \Delta CT_{\text{common}}$ (note that we applied the correction factor for difference in amplification efficiency calculated in Extended Data Fig. 1). To show fold changes normalized to the control siRNA-transfected samples the following equation was used: normalized $\Delta\Delta\Delta CT = \Delta\Delta CT_{\text{average target siRNA}} - \Delta\Delta CT_{\text{average of control siRNA}}$. Then the decrease (−) or increase (+) in dPAS usage was calculated as $\pm 2^{\text{normalized } \Delta\Delta\Delta CT}$.

Oligonucleotides used for qRT–PCR. Cyclin D1 common forward, 5′-CTGCCAGGAGCAGATCGAAG; reverse, 5′-AATGCTCCGGAGAGGAGGGACT; distal forward, 5′-ATCGAGAGGCCAAAGGCT; reverse, 5′-CGTCTTTTTGTCTTCTGCTGGA; DICER1 common forward, 5′-CTCATTATGACTTGCTATGTCGCCTTG; reverse, 5′-CACAATCTCACATGGCTGAGAAG; distal forward 5′-TGCTTTCCGCAGTCCTAACTATG; reverse, 5′-AATGCCACAGACAAAAATGACC; TIMP2 common forward, 5′-CAACCCTATCAAGAGGATCCAGTAT; reverse, 5′-GATGTCGAGAAACTCCTGCTTG; distal forward, 5′-GACATCAGCTGTAATCATTCCTGTG; reverse, 5′-CGATGCCAAATGGAGAGC; FHL1 common forward, 5′-CTGGCACAAAGACTGCTTCACCTGT; reverse 5′-GATTGTCCTTCATAGGCCACCACACTGG; distal forward, 5′-GCCAGGGCTGTCATCAACATGGATA; reverse 5′-TGCATTTCAGGTAAGCGGTAGGTGGA; tubulin common forward, 5′- GAAGGCCTCATCCTCCACTTTGGAAAG; reverse, 5′-TGCTAGCAGTGTCTCATGCTCG; distal forward, 5′-GCATCAGTAGCTGAGTGCACTCCTGGT; reverse, 5′-GTAGAGGGTATGAAGGGCAAGAACTCT; VMA21 common forward, 5′- GATAAGGCGGCGCTGAACGCACTGC; reverse, 5′-TGAGCCTTCATTCCAGGCCACATACACA; distal forward, 5′-CATCTGCACAGCACCTTACAGTTTGC; reverse, 5′-GAAATGCAGCACATCCAAATCCTCCC; GSK-3β common forward, 5′-CTGGTCCGAGGAGAACCCAATGTTTCG; reverse, 5′-CAGCCAACACACAGCCAGCAGACCATAC; distal forward, 5′-GAGCTGAGCCCATGGTTGTGTGTAAC; reverse, 5′-GGTTCACTTCAGCAGGCAGGACAACTC; SMOC1 common forward, 5′-CTCTGATGGCAGGTCCTACGAGTCCA; reverse, 5′-GTATGGCACTGCACCTGGGTAAAGGAG; distal forward, 5′-GAGTCCTGCAATTGTACTGCGGACTCCA; reverse 5′-CATGGGATCTGGACTCCCTTCCTCTC; ACER3 common forward, 5′-CACGCTGGACTGGTGCGAGGAGAACT; reverse, 5′-GTGGAAGCACCAGGATCCCATTCCTACC; distal forward, 5′-CTGTTCAAGCTAATACAGCATTTCCT; reverse, 5′-GTGAATAAGCAGACTGAGATTACCTG; TMEM48 common forward, 5′-CATTCATCCTCAGCAACTCATGCACTC; reverse 5′-CTGTTAGTACCAGTGCAGGGAACCAC; distal forward, 5′-GTGCTGTGTACTAAATACAGGCCACATAGTG; reverse 5′-CCTGGTTCCAACAGATGGTGTGTAGA; MSRB3 common forward, 5′-CTCTGGGAAGTGCGCAGTCCGGGT; reverse, 5′-GTCCCTTTCTCCTGAGTGACATGG; distal forward, 5′-GCAGGATATGGAGTGCAATGAACTGAG; reverse, 5′-ACAGTAAGAGCTGGAGTGCAGAGA; 7SK forward, 5′-GACATCTGTCACCCCATTGATC; reverse, 5′-TCTGCAGTCTTGGAAGCTTGAC.

Luciferase assays. One day after a second hit with siRNA (as described earlier), HeLa cells were transfected with 0.25 μg of gene-specific 3′ UTR *Renilla* luciferase plasmids (SMOC1 and GAPDH from Switchgear Genomics) using Lipofectamine

2000 (Invitrogen). *Renilla* luciferase activity was assayed 24 h after plasmid transfection using Stop and Glo reagent (Promega).

Generation of stable cell lines. LN229 cells were transfected with CFIm25-specific shRNA or control shRNA using polybrene in 6-well plates. Two days after lentiviral transfection cells were transfected with a second hit of lentivirus. Selection was done using 1 $\mu g\,ml^{-1}$ of puromycin over 2 weeks. U251 cells were transfected with either GFP or CFIm25 expressing pcDNA3 plasmids using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Selection was performed over 1–2 weeks using 2.5 $mg\,ml^{-1}$ of G418.

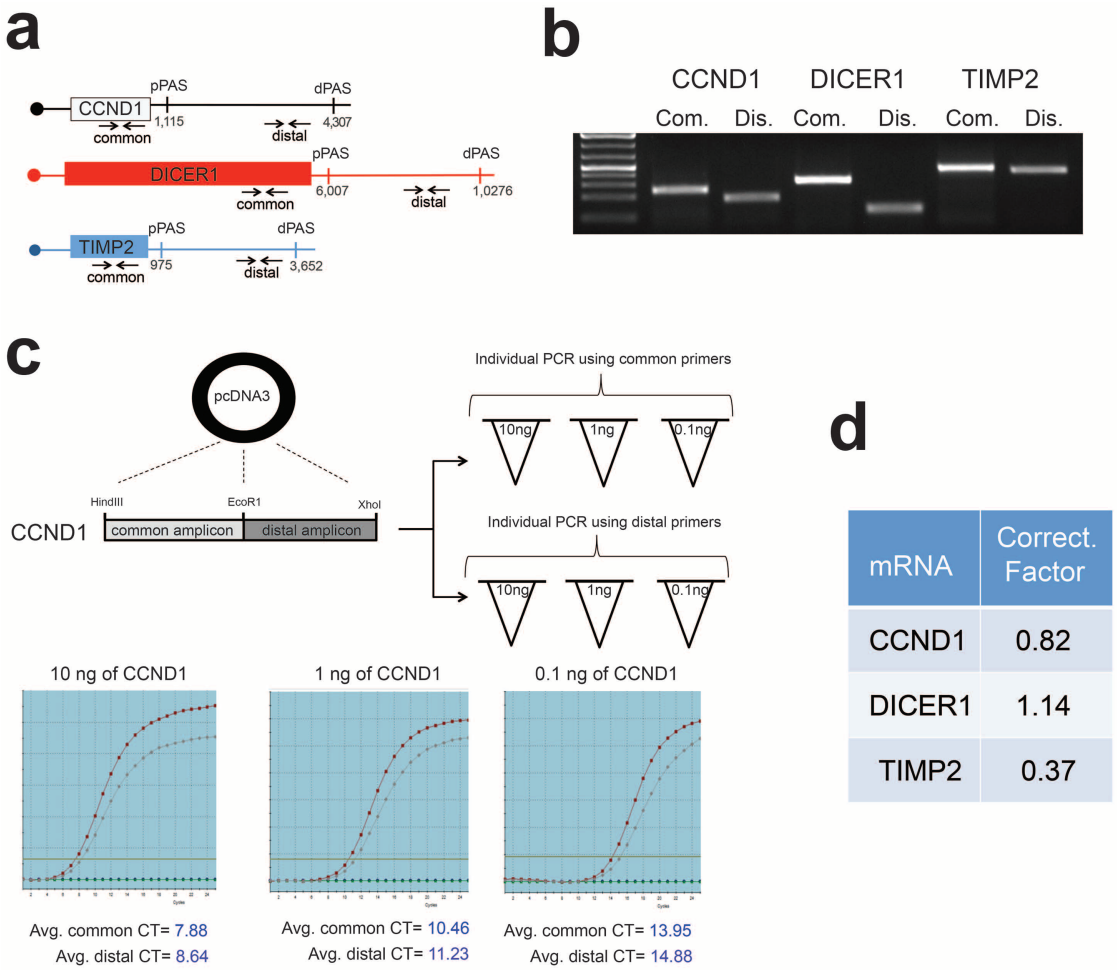
Soft agar assay. Soft agar assays were used to determine anchorage-dependent growth. For the base layer, 1% of UltraPure low melting point agarose (Invitrogen) was mixed 1:1 with 2× DMEM media and plated in 6-well plates giving a 1.5 ml bottom layer of 0.5% agar. Then 3×10^4 cells of LN229 shRNA stably transfected cells were titrated into 2× DMEM and mixed with an equal volume of 0.6% agar to give a 0.3% layer and 1.5 ml was dispensed into each well. The agar was covered with 1 ml of 1× DMEM and incubated in a humidified incubator at 37 °C (5% CO_2). Fresh media was added once a week. After 2 weeks, colonies formed were stained with 0.01% crystal violet, photographed and counted. For U251 plasmid transfected cells the same protocol was followed except that a third (0.3%) layer of agar was plated on top of the layer containing the cell suspension.

Matrigel invasion assay. The Matrigel invasion assay was performed following the manufacturer's protocol. Briefly, the 6-well BioCoat Matrigel Invasion Chamber (Becton Dickinson) was rehydrated with FBS free DMEM. The Matrigel trans-well inserts were then transferred to 6-well plates containing 10% FBS on the bottom. U251 siRNA-transfected or LN229 shRNA-transfected cells were plated (5×10^5 cells per well) in triplicate wells of the upper chamber in serum-free media. After 24 h, cells were stained with 0.01% crystal violet, and the number of invading cells was counted at ×20 magnification in 10 fields for each well.

Statistical tests. Unless otherwise specified, experiments were done using three biological replicates and data are shown as average ± s.d., and statistical analysis was done using a two-tailed student *t*-test.

Subcutaneous xenograft tumour model. Hsd:Athymic Nude-Foxn1nu nude mice at age 5–6 weeks were used. For each cell line (LN229 or U251), 20 male nude mice were randomly assigned into two groups ($n = 10$). Stably transfected LN229 and U251 cells were resuspended in pure culture medium with the concentration of 3×10^7 cells ml^{-1}. One-hundred-microlitre cell suspensions (3×10^6 cells) were inoculated subcutaneously into the lower right flank of the mice using a 27-gauge needle. Tumour diameters are measured with digital callipers, and the tumour volume in mm^3 is calculated by the formula: volume = $(\text{width})^2 \times \text{length}/2$. The tumour size data were collected and processed blindly. The animal experiments were performed under the Institutional Review Board approved animal protocol AWC-13-115.

31. Ward, A. & Dutton, J. R. Regulation of the Wilms' tumour suppressor (*WT1*) gene by an antisense RNA: a link with genomic imprinting? *J. Pathol.* **185,** 342–344 (1998).
32. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
33. Kaplan, P. J., Mohan, S., Cohen, P., Foster, B. A. & Greenberg, N. M. The insulin-like growth factor axis and prostate cancer: lessons from the transgenic adenocarcinoma of mouse prostate (TRAMP) model. *Cancer Res.* **59,** 2203–2209 (1999).
34. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79,** 351–379 (2010).
35. Braulke, T., Dittmer, F., Gotz, W. & von Figura, K. Alteration in pancreatic immunoreactivity of insulin-like growth factor (IGF)-binding protein (IGFBP)-6 and in intracellular degradation of IGFBP-3 in fibroblasts of IGF-II receptor/IGF-II-deficient mice. *Horm. Metab. Res.* **31,** 235–241 (1999).
36. Hu, J. F. *et al.* Lack of reciprocal genomic imprinting of sense and antisense RNA of mouse insulin-like growth factor II receptor in the central nervous system. *Biochem. Biophys. Res. Commun.* **257,** 604–608 (1999).
37. Ellis, M. J. *et al.* Insulin-like growth factors in human breast cancer. *Breast Cancer Res. Treat.* **52,** 175–184 (1998).
38. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92–105 (2009).
39. Gómez-Angelats, M., Teeguarden, J. G., Dragan, Y. P. & Pitot, H. C. Mutational analysis of three tumor suppressor genes in two models of rat hepatocarcinogenesis. *Mol. Carcinog.* **25,** 157–163 (1999).

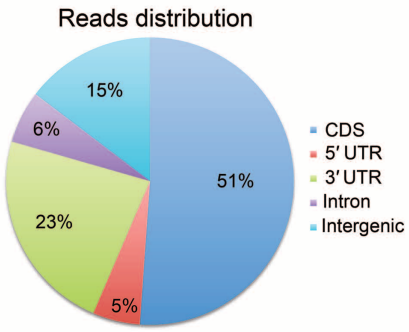


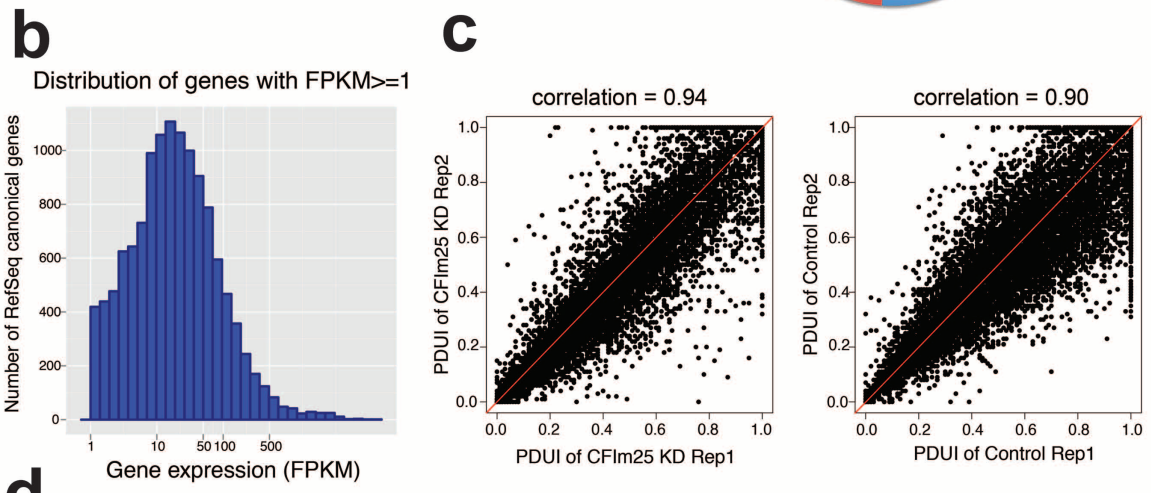
mRNA	Correct. Factor
CCND1	0.82
DICER1	1.14
TIMP2	0.37

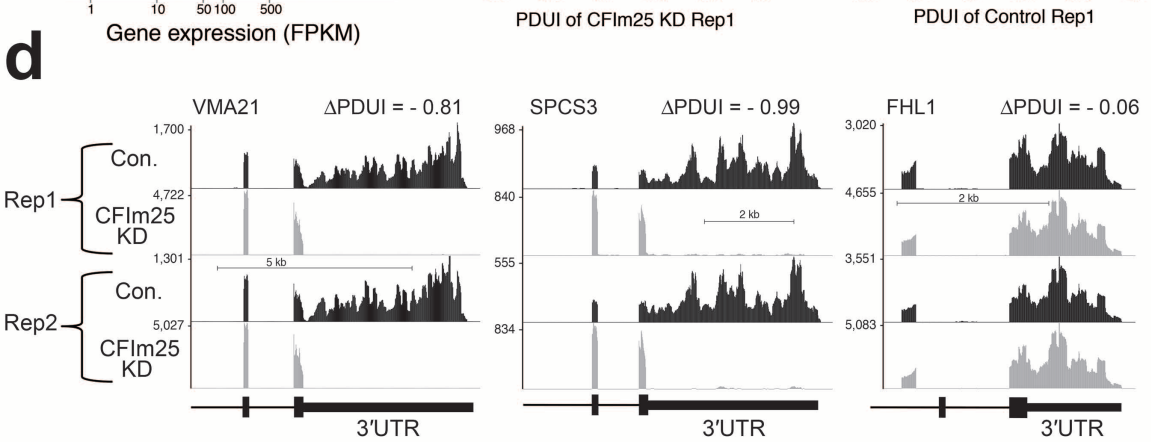
Extended Data Figure 1 | Design and optimization of the qRT–PCR assay to monitor APA of three test genes. **a**, Schematic denotes the relative location of the common and distal primer annealing sites in each test gene and the approximate locations of the annotated proximal and distal poly(A) sites, depicted as pPAS and dPAS, respectively. The numbers demarcate where the 3′ UTR starts and ends according to ENSEMBL. **b**, Ethidium-stained agarose gel of RT–PCR products of equal cycle number from the different amplicons using HeLa cell mRNA. **c**, Both the common and distal cyclin D1 amplicons were cloned into the same pcDNA3 plasmid in tandem. Three dilutions of each plasmid were made and amplified individually with each amplicon in triplicate. The two lines on the graph depict the amplification curve for the common and distal amplicons. The expectation is that identical cycle threshold (CT) values should be attained for each, given that the PCR reactions were conducted using identical amounts of starting material. The average of three individual experiments is shown for each dilution and the average CT deviation of either amplicon at all of the dilutions was calculated as a correction factor. **d**, The experiment shown in **c** was repeated for DICER1 and TIMP2 to determine their respective correction factors, which was then applied to experiments shown in Fig. 1.

a

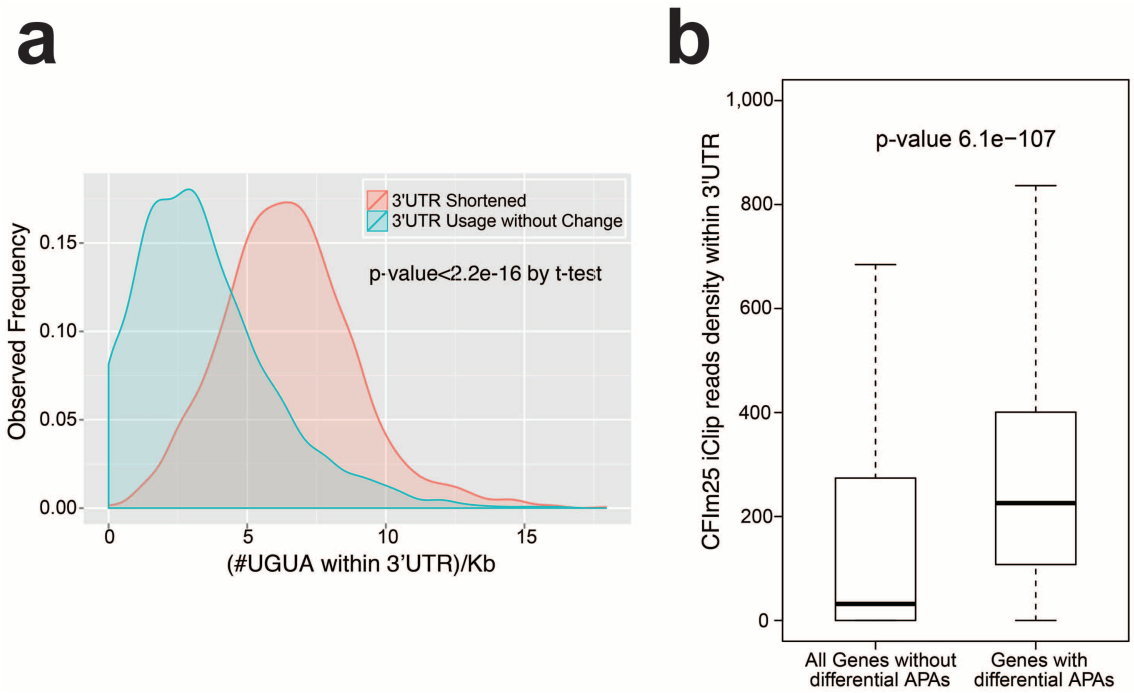
Samples	Con. R1	Con. R2	CFIm25KDR1	CFIm25KD R2
Mapped reads	269,463,902	270,426,980	265,270,708	282,404,060
#CDS	52.4%	47.6%	53.0%	51.5%
#5' UTR	5.2%	5.2%	5.5%	5.4%
#3' UTR	27.0%	23.4%	21.4%	20.3%
#Intron	6.3%	5.8%	6.0%	5.6%
#Intergenic	9.1%	17.9%	14.1%	17.3%
3' UTR average coverage	236	205	184	185



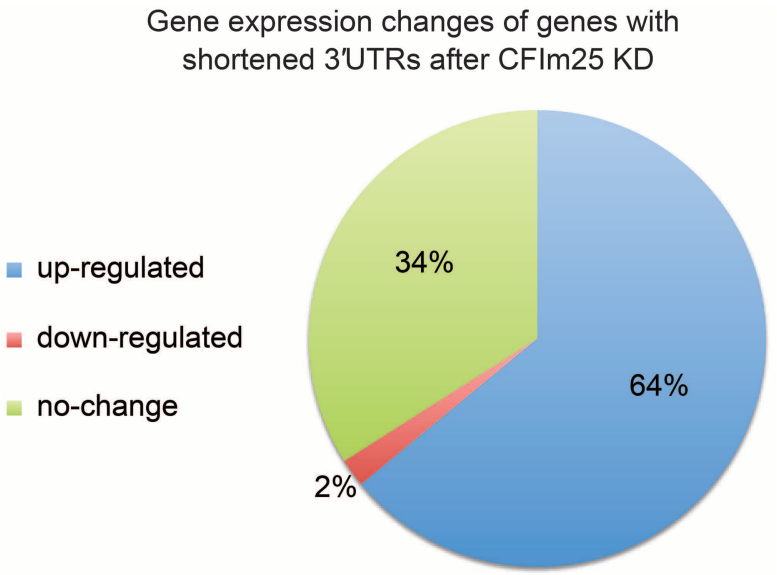




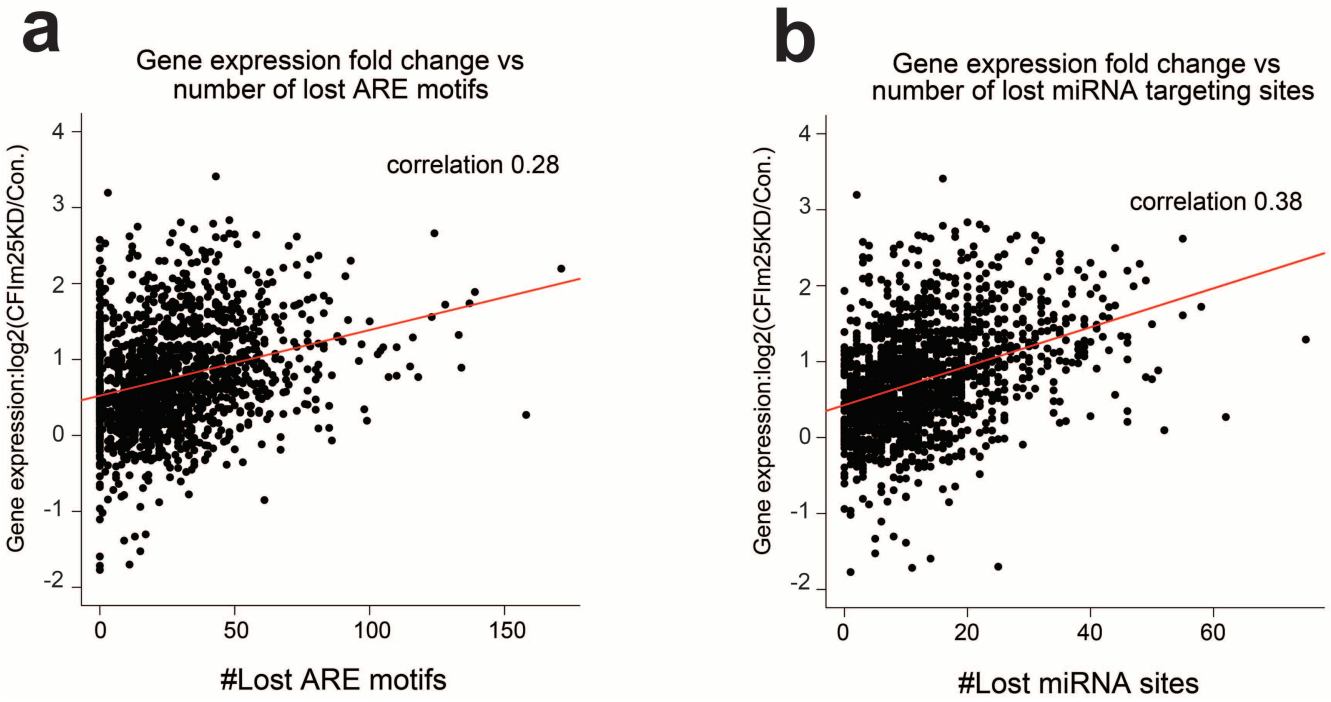
Extended Data Figure 2 | Summary of RNA-seq alignment and reproducibility of PDUI and CFIm25-depletion-induced 3′ UTR shortening. a, RNA-seq read statistics of the four biologically independent experiments where HeLa cells were treated with either control siRNA (Control) or CFIm25 siRNA (CFIm25kD). Pie chart on the right represents genomic distribution of reads that were mapped to human genome hg19. The percentage was calculated by averaging all samples. CDS, coding region. **b**, Histogram of gene expression of RefSeq genes with fragments per kilobase of transcript sequence per million mapped paired-end reads (FPKM) no less than 1. **c**, Scatterplot of the two biological replicates for each condition with high Pearson correlation ($r \geq 0.9$) demonstrating a high level of reproducibility between sample PDUI scores. Each dot represents the PDUI of a transcript. **d**, Genome browser screen images from four independent RNA-seq experiments. Each represents an independent biological sample where HeLa cells were transfected with either the control siRNA (Con.) or an siRNA that knocked down CFIm25. Both VMA21 and SPCS3 were found to undergo 3′ UTR shortening after CFIm25 knockdown whereas FHL1 was found not to change.



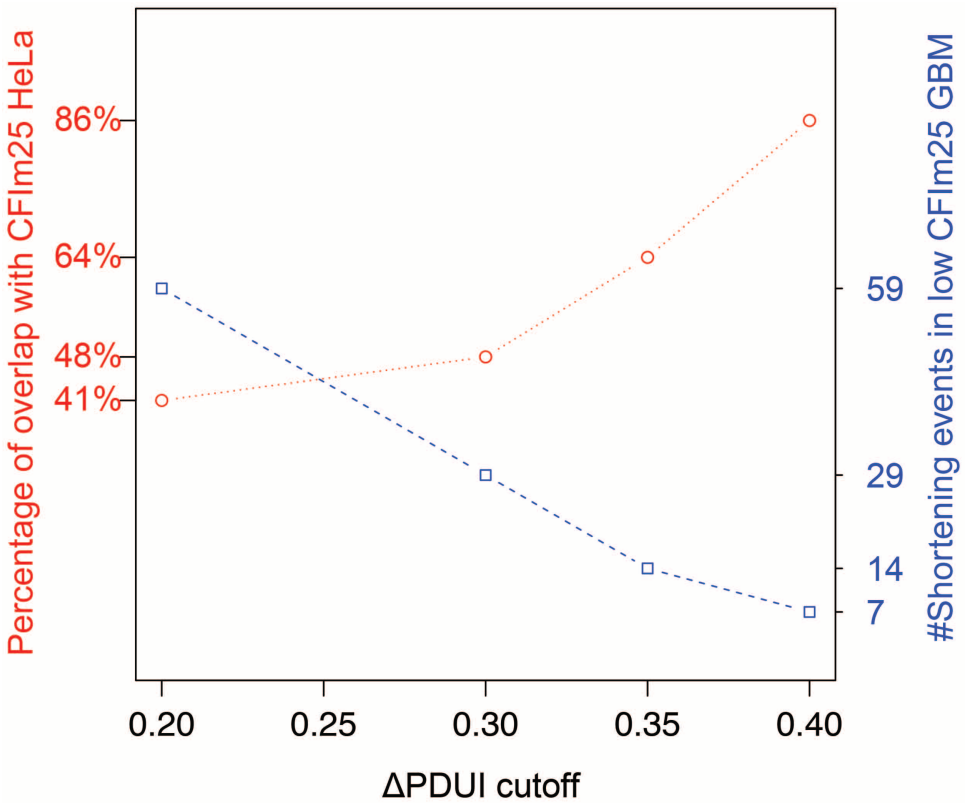
Extended Data Figure 3 | Shortened transcripts have more UGUA CFIm25-binding motifs than unaltered transcripts. **a**, CFIm25 is known to bind to the UGUA motif. The number of UGUA motifs within the 3′ UTRs of genes with 3′ UTR shortening after CFIm25 knockdown relative to genes with unaltered 3′ UTRs was calculated and compared. Here we selected the genes without 3′ UTR change according to them having a ΔPDUI value ≤ 0.05. **b**, iCLIP tags from ref. 14 (Gene Expression Omnibus accession number GSE37398) were superimposed onto data derived from PDUI analysis of CFIm25 knockdown cells. The box plot demonstrates the enrichment of CFIm25 binding within 3′ UTRs that are altered after CFIm25 knockdown ($P = 6.1 \times 10^{-107}$, *t*-test).



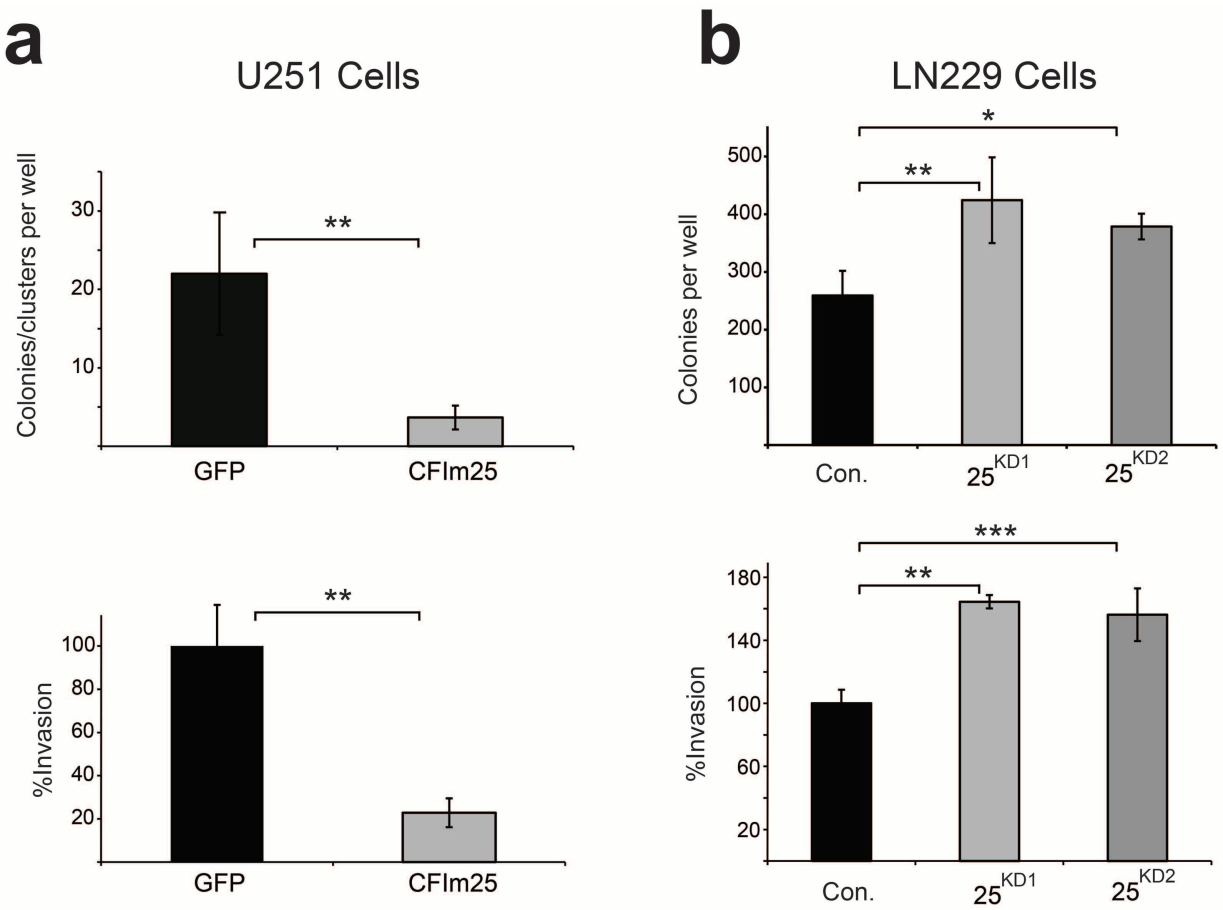
Extended Data Figure 4 | Gene expression changes of genes with shortened 3′ UTRs. Pie chart was calculated from the list of 1,450 genes exhibiting shortened 3′ UTRs due to CFIm25 knockdown (dn, down). Differentially expressed gene analysis was performed using edgeR with FDR ≤ 0.05 (see Methods).



Extended Data Figure 5 | The Pearson correlation between gene expression fold change and the number of lost negative regulatory elements. Left, the number of lost AREs (AU-rich elements) due to 3′ UTR shortening was calculated using the ARE database and plotted against change in gene expression levels after CFIm25 knockdown (KD). Right, similar to the left except the number of lost patented miRNA target sites (Targetscan 6.2) was plotted.



Extended Data Figure 6 | Overlap between shortening events in glioblastoma with low CFIm25 and shortening events in HeLa cells after CFIm25 knockdown. Left, *y*-axis (red) represents the percentage of shortening events in low CFIm25 glioblastoma that are also shortened in HeLa cells after CFIm25 knockdown. Right, *y*-axis (blue) shows the number of shortening events in low CFIm25 glioblastoma (GBM) against different ΔPDUI cut-offs.



Extended Data Figure 7 | Overexpression of CFIm25 reduces invasion and colony formation whereas CFIm25 depletion increases invasion and colony formation. **a**, U251 cells were transfected with either GFP or CFIm25. Top left, Cells were replated in soft agar and the number of colonies/clusters formed were determined. Bottom left, Matrigel invasion assay for cells overexpressing CFIm25 or GFP. **b**, Top right, LN229 cells were transfected with either control or two different lentiviral plasmids targeting CFIm25 (KD1 and KD2). Stably transfected cells were plated on soft agar and the resulting colonies were counted for KD1 and KD2, respectively. Bottom right, LN229 cells were transfected with either control or two different siRNAs (KD1 and KD2) directed against CFIm25 and were replated for a Matrigel invasion assay. All the experiments were done in biological triplicates and shown is the mean ± s.d. All *P* values were from the two-tailed student *t*-test of the control versus sample. $*P < 0.1$, $**P < 0.01$, $***P < 0.001$.

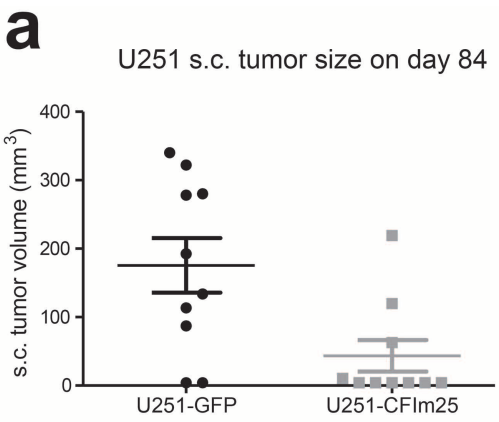


Table Analyzed	U251 s.c. tumor size on day 84
Column A	U251-GFP
vs	vs
Column B	U251-CFIm25
Unpaired t test	
P value	0.0102
P value summary	*
Are means signif. different? (P < 0.05)	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.869 df=18

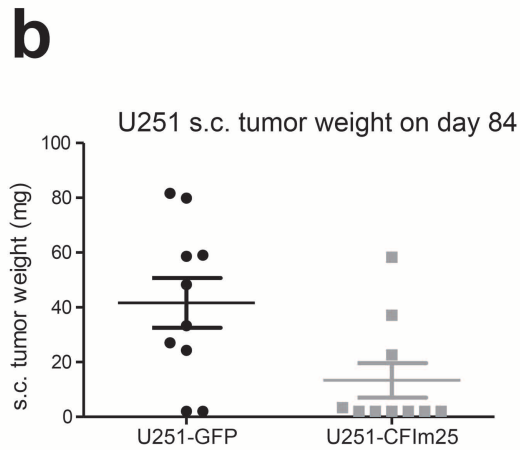
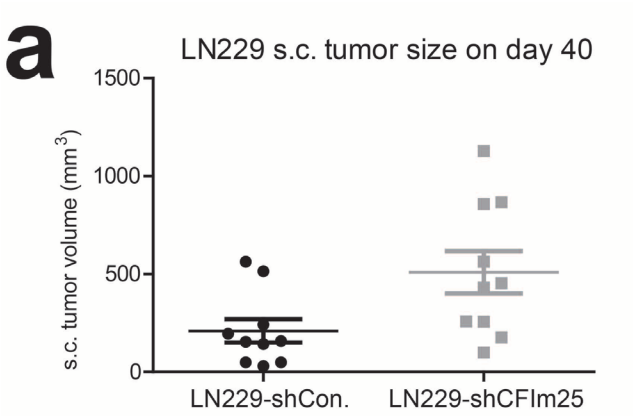
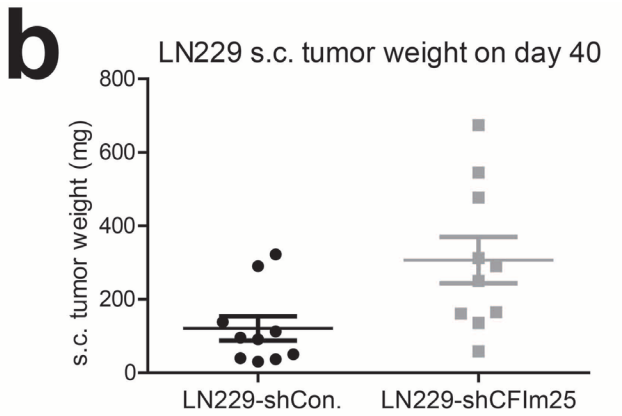


Table Analyzed	U251 s.c. tumor weight on day 84
Column A	U251-GFP
vs	vs
Column B	U251-CFIm25
Unpaired t test	
P value	0.0197
P value summary	*
Are means signif. different? (P < 0.05)	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.560 df=18

Extended Data Figure 8 | Overexpression of CFIm25 in U251 tumours reduces their size and weight. **a**, **b**, U251 subcutaneous (s.c.) xenograft tumours were isolated from nude mice on day 84 after implantation and measured for volume (**a**) and weight (**b**) ($n = 10$). U251-GFP indicates control U251 cells expressing GFP and U251-CFIm25 indicates cells transduced with a lentivirus that overexpresses CFIm25.



Unpaired t test	
P value	0.0257
P value summary	*
Are means signif. different? (P < 0.05)	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.431 df=18



Unpaired t test	
P value	0.0173
P value summary	*
Are means signif. different? (P < 0.05)	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.621 df=18

Extended Data Figure 9 | Reduction in CFIm25 expression levels enhances LN229 tumour size and weight. **a**, **b**, LN229 subcutaneous (s.c.) xenograft tumours were isolated from nude mice on day 40 after implantation and measured for volume (**a**) and weight (**b**) ($n = 10$). LN229-shCon. indicates control lentiviral transduced cells and LN229-shCFIm25 indicates cells transduced with a lentivirus that expresses shRNA targeting CFIm25.